DYNAMICS OF BOUNCING RIGID BODIES AND BILLIARDS IN THE SPACES OF CONSTANT CURVATURE

BY

KI YEUN KIM

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Mathematics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2016

Urbana, Illinois

Doctoral Committee:

Professor Yuliy Baryshnikov, Chair
Professor Vadim Zharnitsky, Director of Research
Associate Professor Robert DeVille
Associate Professor Zoi Rapti

Abstract

Mathematical billiard is a dynamical system studying the motion of a mass point inside a domain. The point moves along a straight line in the domain and makes specular reflections at the boundary. The theory of billiards has developed extensively for itself and for further applications. For example, billiards serve as natural models to many systems involving elastic collisions. One notable example is the system of spherical gas particles, which can be described as a billiard on a higher dimensional space with a semi-dispersing boundary.

In the first part of this dissertation, we study the collisions of a two-dimensional rigid body using billiard dynamics. We first define a dumbbell system, which consists of two point masses connected by a weightless rod. We assume the dumbbell moves freely in the air and makes elastic collisions at a flat boundary. For arbitrary mass choices, we use billiard techniques to find the sharp bound on the number of collisions of the dumbbell system in terms of the mass ratio. In the limiting case where the mass ratio is large and the dumbbell rotates fast, we prove that the system has an adiabatic invariant. In case the two masses of the dumbbell are equal, we assume gravity in the system and study its infinitely many collisions. In particular, we analytically verify that a Smale horseshoe structure is embedded in the billiard map arising from the equal-mass dumbbell system.

The second part of this dissertation concerns the billiards in the spaces of constant curvature. We provide a unified proof to classify the sets of three-period orbits in billiards on the Euclidean plane, the hyperbolic plane and on the two-dimensional sphere. We find that the set of three-period orbits in billiards on the hyperbolic plane has zero measure. For the sphere, three-period orbits can form a set of positive measure if and only if a certain natural condition on the orbit length is satisfied.

To my parents who believed in me and supported me all along.

Acknowledgments

I would like to express my gratitude to my advisor Vadim Zharnitsky for his consistent guidance and encouragement. I would also like to thank the rest of my committee members, Yuliy Baryshnikov, Lee DeVille and Zoi Rapti, for their helpful comments and insights.

This work was financially supported by Dr. Lois M. Lackner mathematics fellowship, the Hone-Nash award, and a gift to the Mathematics department at the University of Illinois from Gene H. Golub. I deeply appreciate the generosity of the donors.

I am grateful to my family and friends who have supported me through my ups and downs during the past years. A special thanks goes to Mike for proof-reading my work as a fellow mathematician and as a native speaker, and for being a great part of my life.

Table of Contents

CHAPTER 1

Introduction

Mathematical billiards is the study of the free motion of a mass point in a domain: the point moves along a "straight line" inside the domain, and makes mirror-like reflections at the boundary. These billiard models are simple yet they illustrate various fundamental concepts in dynamical systems. Moreover, billiards have close connections to other fields of mathematics as well as to other sciences. In this dissertation, we focus on mathematical billiards and its applications to the motion of a simple rigid body. Along the way, we explore topics from classical mechanics and chaos theory.

The plan of the paper is the following. In the remainder of this introduction, we briefly explain our main results and provide some motivation to the study. In Chapter 2, we discuss essential terminologies and other relevant background materials in detail. In Chapter 3, we introduce a *dumbbell* system consisting of two point masses. We present the main results regarding the dumbbell in Chapter 4 through Chapter 6. Chapter 4 concerns the dumbbell with two arbitrary masses, whereas Chapter 5 considers the case of a limiting mass ratio and Chapter 6 considers the case of equal masses with an extra assumption of the gravitational force. Chapter 7 contains independent topics and results about periodic orbits in billiards on the non-Euclidean spaces.

Let us begin with a motivation why we study the dumbbell. Many mechanical systems which involve elastic collisions are often easier to analyze when they are considered as billiard systems. In the 1960s, the billiard problem appeared in the context of statistical mechanics to study a system of spherical gas particles to verify the Boltzmann ergodic hypothesis [Si]. More recently, non-spherical particles, which have rotational in addition to translational velocities, were studied to build a more realistic gas model [Co]. In this direc-

tion, it is natural to consider the behavior of "a billiard ball with two degrees of freedom" moving inside a domain. Even in the case where the boundary is flat, the extra degree of freedom of a moving object can generate many interesting phenomena including *deterministic chaos* and *adiabatic invariance*.

In Chapter 3, we introduce a dumbbell system, which consists of two point masses m_{L} and m_{R} connected by a weightless rigid rod. We study the elastic collisions of the dumbbell with a flat boundary. This is an important first step before understanding the full dynamics of a two-dimensional object in some general planar domains.

If we assume no gravity, the dumbbell will leave the flat boundary after finitely many collisions. In Chapter 4, by first considering the dumbbell system as a single point billiard then applying the "*unfolding technique*" [Ta2], we estimate the maximal number of collisions of the dumbbell in terms of the mass ratio $\frac{m_{\mathrm{L}}}{m_{\mathrm{R}}}$ [Ba2]. One of the basic questions about an ideal gas model is to estimate the maximum number of collisions for a system of spherical gas particles moving freely in an open space [Bu]. Our collision estimate of the dumbbell can be useful to extend such study to a system of non-spherical gas particles.

For the previous result, we did not impose any assumptions on the two masses m_{L} and m_{R}. We now consider the two natural special cases when the mass ratio $\frac{m_{\mathrm{L}}}{m_{\mathrm{R}}} \ll 1$ or when $\frac{m_{\mathrm{L}}}{m_{\mathrm{R}}} = 1$. In the former case, we prove that there exists a function, which depends on the phase variables of the dumbbell system and a slowly varying parameter, such that the function value remains "nearly constant" along an orbit. We call this function an *adiabatic invariant*.

One classical example of adiabatic invariance is in the Fermi-Ulam model. It is a one-dimensional system where an elastic particle travels between two slowly moving walls. One can use the laws of physics to show that the product of the velocity of the particle and the distance between the walls is an adiabatic invariant. In Chapter 5, we show that the dumbbell system, which is a two-dimensional system, has a comparable adiabatic invariant. If we assume the dumbbell has a small mass ratio $\frac{m_{\mathrm{L}}}{m_{\mathrm{R}}} \ll 1$ and approaches the flat boundary very slowly, then the adiabatic invariant of the system is given by the product of the angular velocity of the dumbbell and the angular distance traveled by the lighter mass m_{L} between

two consecutive collisions.

For the second case when the two masses are equal, we add gravity to the system and interpret this scenario of the equal-mass dumbbell with gravity as a special case of a coin toss with collisions. Previous work on the dynamics of a coin either assumes a tossed coin is caught in hand before a collision [Di, Ke] or investigates the collision effects of a coin using numerical simulations [Str, Vu]. Here, we present our analytic result on the collision dynamics of a two-dimensional coin: a thin disk whose motion is restricted in the two-dimensional space.

Note that our equal-mass dumbbell can be viewed as an equivalent model to the two-dimensional coin. That is, despite having the different moments of inertia, the basic motions of the equal-mass dumbbell and the two-dimensional coin are the same. Since we already have established the billiard dynamics for the dumbbell, we continue to use the dumbbell to study the two-dimensional coin. Our results below, which are proven using the equal-mass dumbbell, remain true if we use the coin instead.

The main results in Chapter 6 are on the chaotic behavior of the two-dimensional coin (equal-mass dumbbell). Since the gravity pulls the coin back to the boundary and there is no energy loss, the coin will bounce infinitely many times. We prove that the two-dimensional bouncing coin can produce any infinite sequence of collisions: if a collision of m_{L} or m_{R} is labeled by L or R respectively, then any infinite collision sequence of L's and R's can be realized by choosing an appropriate initial condition.

The proof begins by viewing the coin system as a single point gravitational billiard in a transformed domain. Due to gravity, the mass point moves along a parabola inside the domain. Then we show that the *Smale horseshoe* [Sm] is embedded in the billiard map.

Motivated from the study of a nonlinear forced oscillator, the Smale horseshoe map is one of the famous examples of a chaotic map. The horseshoe map is defined topologically: in the standard setting, it transforms a square into a long and thin strip and places the strip over the square in the shape of a horseshoe. When iterated forward and backward infinitely many times, the horseshoe map generates an invariant Cantor set in the square. The horseshoe map acting on the invariant set is topologically conjugate to a shift map on

the set of infinite sequences of two symbols. Therefore, once we have a horseshoe, we can use the fact that there exists an one-to-one correspondence between an infinite symbolic sequence and a phase point to complete the proof.

One way to verify that a certain map contains a horseshoe is by showing that the map satisfies the Conley-Moser conditions [Mo, Wi], which are a combination of geometric and analytic criteria. If met, these conditions guarantee the existence of a horseshoe.

Although there are many systems in science and engineering that are proven to contain a horseshoe [Lv, Ho], to the best of our knowledge, all the already-known horseshoe maps are defined on one connected domain. For our coin system, we iterate the billiard map on the union of two disjoint rectangles. We prove that the billiard map on the two rectangles produces the horizontal and vertical strips satisfying the Conley-Moser conditions. The associated set of sequences contains combinatorially more complex sequences than the sequences for the standard horseshoe map.

These results on the bouncing coin assert the existence of chaotic behavior in the system, which leaves the question of how the entire phase space of the coin system will evolve over time. Since the system is highly nonlinear, only partial numerical results are known [Mi, Stz]. The billiard boundary for the coin consists of piecewise-smooth dispersing curves forming convex wedges of angle $\frac{\pi}{2}$ at the discontinuities. It is known that a gravitational billiard in a straight wedge of angle bigger than $\frac{\pi}{2}$ is ergodic [Le, Wo]. Also, when a gravitational force does not exist, all dispersing billiards are ergodic [Si]. Since the billiard system for the coin has a boundary which is some sort of a combination of these, we expect that the coin system is ergodic. We plan to pursue this question as future work.

In Chapter 7, we turn our attention to billiards on the hyperbolic plane and on the 2-sphere. One important question in billiards concerns the presence of large sets of periodic orbits. Ivrii conjectured [Iv] that, in every bounded planar billiard domain with a smooth boundary, the set of n-period orbits has measure zero for all n. The Weyl estimate formula, which describes eigenvalue asymptotics for the Dirichlet problem, has a certain special form if the set of periodic orbits in the associated billiard has measure zero.

The conjecture has been proven for $2, 3$ and 4-period orbits only. Since 2-period orbits

must be normal to the boundary at both ends, we can easily see that in any billiard the set of 2-period orbits has measure zero. The conjecture for the set of $3, 4$-period orbits was proved in various literature [Gl, Gl2, Ry, Sto, Wo2]. There have been extensions to other types of billiard systems such as higher dimensional [Vo], outer [Ge, Tu], and spherical billiards [Ba3].

We present a unified approach to study the set of 3-period orbits for the spherical and the hyperbolic cases [Bl]. We extend the Jacobi fields approach from Wojtkowski's paper [Wo2] to write a (underlying geometry based) relation which must be satisfied on a neighborhood filled with 3-period orbits. On the hyperbolic plane, we obtain a contradiction in the relation and conclude that there cannot be an open set of 3-period orbits. For the sphere, the relation remains true if and only if a certain natural condition on the orbit length is satisfied. Thus, under the condition, the set of 3-period orbit can have a positive measure. Otherwise, it has an empty interior and has measure zero.

It is expected that understanding the structure of the sets of periodic orbits in non-Euclidean billiards would help studying eigenvalue asymptotics in non-Euclidean geometries.

CHAPTER 2

Preliminaries

Before we move on to the main results, we provide some necessary background materials. In Chapter 2.1, we exposit the basics of the Lagrangian and the Hamiltonian mechanics and explain the notion of adiabatic invariance. Many definitions and results in this chapter come from the book by Arnold [Ar]. In Chapter 2.2, we first discuss deterministic chaos, then give the definition of the Smale horseshoe map and the sufficient criteria for the existence of the map. The book by Wiggins is an excellent reference for this topic [Wi]. Finally in Chapter 2.3, we talk about mathematical billiards. For more details about billiards, we recommend the books by Tabachnikov [Ta, Ta2] and the book by Cvitanovic [Cv].

2.1 Lagrangian and Hamiltonian mechanics

2.1.1 Lagrange's equations

Newtonian mechanics is one branch of classical mechanics, and is indisputably a breakthrough development to study the motion of bodies. However, it mainly uses Cartesian coordinates which may lead to extremely complex calculations depending on the systems. Lagrangian mechanics is a reformulation of Newtonian mechanics, which is better adapted to the change of coordinates.

The *configuration space* is the the collection of the parameters that defines the position of a system. These parameters which we denote by $\mathbf{q} = (q_1, ...q_n)$ are called the *generalized coordinates*. The time derivatives of the generalized coordinates, $\dot{\mathbf{q}}$, are called the *generalized velocities*. In general, a configuration space is a differential manifold.

For a conservative mechanical system on its configuration space, the *Lagrangian* is defined by $L(\mathbf{q}, \dot{\mathbf{q}}, t) = K - U$ where K is the kinetic energy and U is the potential energy of the system.

Lagrangian mechanics is based on the calculus of variation which concerns the extremals of functionals: functions from the space of curves to the reals. Given a Lagrangian $L(\mathbf{q}, \dot{\mathbf{q}}, t)$, we call the functional $\Phi(\gamma) = \int_{t_0}^{t_1} L(\mathbf{q}, \dot{\mathbf{q}}, t)dt$ the *action* of the system. The evolution of $\mathbf{q}$ is governed by the following variational condition on the action functional.

Theorem 2.1. *Hamilton's principle of least action states that between two points $\mathbf{q}(t_0)$ and $\mathbf{q}(t_1)$, the motion of the system follows the curve γ such that the action $\Phi(\gamma) = \int_{t_0}^{t_1} L(\mathbf{q}, \dot{\mathbf{q}}, t)dt$ is an extremum.*

Theorem 2.2. *The derivative of the functional $\Phi(\gamma) = \int_{t_0}^{t_1} L(\mathbf{q}, \dot{\mathbf{q}}, t)dt$ is given by*

$$F(h) = \int_{t_0}^{t_1} \Big[\frac{\partial L}{\partial \mathbf{q}} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\mathbf{q}}}\right)\Big] h dt + \left(\frac{\partial L}{\partial \dot{\mathbf{q}}} h\right)\Big|_{t_0}^{t_1}$$

where h is the variation of the curve γ.

An extremal of the action $\Phi(\gamma)$ is a curve such that $F(h)$ is equal to zero. From there, we get Lagrange's equations, which determine the trajectory of $\mathbf{q}$ on the configuration space.

Definition. *Lagrange's equations* for the functional $\Phi = \int_{t_0}^{t_1} L(\mathbf{q}, \dot{\mathbf{q}}, t)dt$ are given by

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\mathbf{q}}}\right) - \frac{\partial L}{\partial \mathbf{q}} = 0.$$

2.1.2 Hamilton's equations

Hamiltonian mechanics is another formulation of classical mechanics, where a system is expressed in terms of a set of canonical coordinates $(\mathbf{q}, \mathbf{p})$. In a mechanical system, $\mathbf{q}$ are the generalized coordinates and $\mathbf{p} = \frac{\partial L}{\partial \dot{\mathbf{q}}}$ are the *generalized momenta*.

We can convert a Lagrangian system to a Hamiltonian system via the Legendre transformation.

Definition. Let $f(x)$ be a convex real-valued function. Given a number p, let

$$x^*(p) = \max_x \{px - f(x)\}.$$

Such $x^*(p)$ can be found by solving $f'(x) = p$. The *Legendre transform* of $f(x)$ is a new function

$$g(p) = px^*(p) - f(x^*(p)).$$

Theorem 2.3. *Suppose Lagrange's equations describe the evolution of* $\mathbf{q}$ *on an* n*-dimensional configuration space. The system of Lagrange's equations is equivalent to the system of* $2n$ *first order differential equations, called Hamilton's equations*

$$\dot{\mathbf{p}} = -\frac{\partial H}{\partial \mathbf{q}} \qquad \dot{\mathbf{q}} = \frac{\partial H}{\partial \mathbf{p}},$$

where $H(\mathbf{p}, \mathbf{q}, t) = \mathbf{p}\dot{\mathbf{q}} - L(\mathbf{q}, \dot{\mathbf{q}}, t)$ *is the Legendre transform of the Lagrangian function viewed as a function of* $\dot{\mathbf{q}}$*. We call* H *the Hamiltonian. In a mechanical system, the Hamiltonian coincides with the total energy of the system.*

The evolution of the system is given by the solution to Hamilton's equations. When interpreted geometrically, it means that the evolution is described as the flow on a vector field on the phase space, a $2n$-dimensional symplectic manifold.

The Hamiltonian formalism allows us to canonically transform one set of canonical coordinates $(\mathbf{p}, \mathbf{q})$ to another set $(\mathbf{P}, \mathbf{Q})$. That is, the transformation conserves the 2-form, $\Sigma\, dp_i \wedge dq_i = \Sigma\, dP_i \wedge dQ_i$. In the next section, we give an example of such transformation.

2.1.3 Adiabatic invariance

Consider a Hamiltonian system with one degree of freedom, which exhibits periodic motion. Suppose the system depends on a slowly varying external parameter λ and $H(p, q; \lambda)$

is a twice differentiable in λ.

$$H = H(p, q; \lambda), \qquad \dot{p} = -\frac{\partial H}{\partial q} \qquad \dot{q} = \frac{\partial H}{\partial p}. \tag{2.1}$$

Then for the system (2.1), there exists a function of the phase variables and the parameter λ, called an *adiabatic invariant* $I(p, q, \lambda)$, such that the value of I remains nearly constant along an orbit.

An adiabatic invariant can be intuitively found as follows. For a fixed λ, the periodicity in motion implies that an orbit on the phase space must be a closed curve. We denote the area bounded by such a curve the *action*. A uniformly varying angular variable on this phase curve is called the *angle* and is usually denoted by φ. As λ slowly changes, the closed phase curve evolves slowly. However, the enclosed area stays nearly constant and it is the adiabatic invariant. A formal definition of an adiabatic invariant is given below.

Definition. The quantity $I(p, q, \lambda)$ is called an *adiabatic invariant* if for every $\kappa > 0$ there exists an $\epsilon_0 > 0$ such that if $0 < \epsilon < \epsilon_0$ and $0 < t < 1/\epsilon$, then

$$|I(p(t), q(t); \epsilon t) - I(p(0), q(0); 0)| < \kappa.$$

Although we did not formally prove here, the transformation of the phase variables (p, q) to the action-angle variables (I, φ) is canonical. It can be done using an indirect method involving a generating function.

If we drop the smoothness condition from the system (2.1), then it does not follow immediately, even in the case of one-dimensional systems, that the action is an adiabatic invariant. In Chapter 5, we consider a non-smooth system with two degrees of freedom and prove that the system has an adiabatic invariant.

2.2 The Smale horseshoe map

2.2.1 Deterministic chaos

In dynamical systems, chaos refers to high sensitivity to initial states in a deterministic system. Although no universal definition of chaos exists, we often determine the chaoticity of a system using *Lyapunov exponents*. Below is a definition of the Lyapunov exponents in a two-dimensional discrete map.

Definition. Let f be a smooth map on $\mathbb{R}^2$. Consider $B(x_0, \delta)$, a disk of small radius δ centered at x_0. Then $Df^n(x_0)(B(x_0, \delta))$ is an ellipse where the two axes have the lengths $r_1^n \geq r_2^n$. These two numbers measure the separation rates of the points near x_0 after n-iterations of f. The Lyapunov numbers of x_0 are defined by

$$\mathcal{L}_1 = \lim_{n\to\infty} \left(\frac{r_1^n}{\delta}\right)^{1/n} \qquad \mathcal{L}_2 = \lim_{n\to\infty} \left(\frac{r_2^n}{\delta}\right)^{1/n}.$$

The *Lyapunov exponents* of x_0 are given by $\ln \mathcal{L}_1$ and $\ln \mathcal{L}_2$.

When the maximal Lyapunov exponent is greater than 0, we generally say the system is chaotic. There are some counter-examples such as the Perron effects where the positive maximal Lyapunov exponent does not indicate chaos.

The history of chaos theory goes back to the time of Poincaré, when he studied the stability of the three-body problem in celestial mechanics. During his study, he introduced a Poincaré section $\mathcal{P}$, a transversal hyperplane to the flow of the system. Then he considered the Poincaré map f defined on $\mathcal{P}$, which takes the intersection of the flow with the section $\mathcal{P}$ to the next intersection. If the flow is periodic, then there exists a fixed point $p \in \mathcal{P}$ such that $f(p) = p$. When $Df(p)$ has two eigenvalues where one is smaller than 1 and one is bigger than 1, we say p is hyperbolic and of a saddle type. Then by the stable manifold theorem, we can find the stable manifold W_S and unstable manifold W_U on which the hyperbolic fixed point p lies. Since by definition $p \in W_S \cap W_U$ implies $f^n(p) \in W_S \cap W_U$, one intersection of the two manifolds imply infinitely many intersections of the two. This

creates a complicated structure which we call the *homoclinic tangle*.

The discovery of a homoclinic tangle is considered as the first observation of chaos. Due to its structure, the iteration of the map separates nearby trajectories arbitrarily far. In the next section, we introduce the Smale horseshoe map, a prototypical map possessing homoclinic invariant points.

2.2.2 The Smale horseshoe map

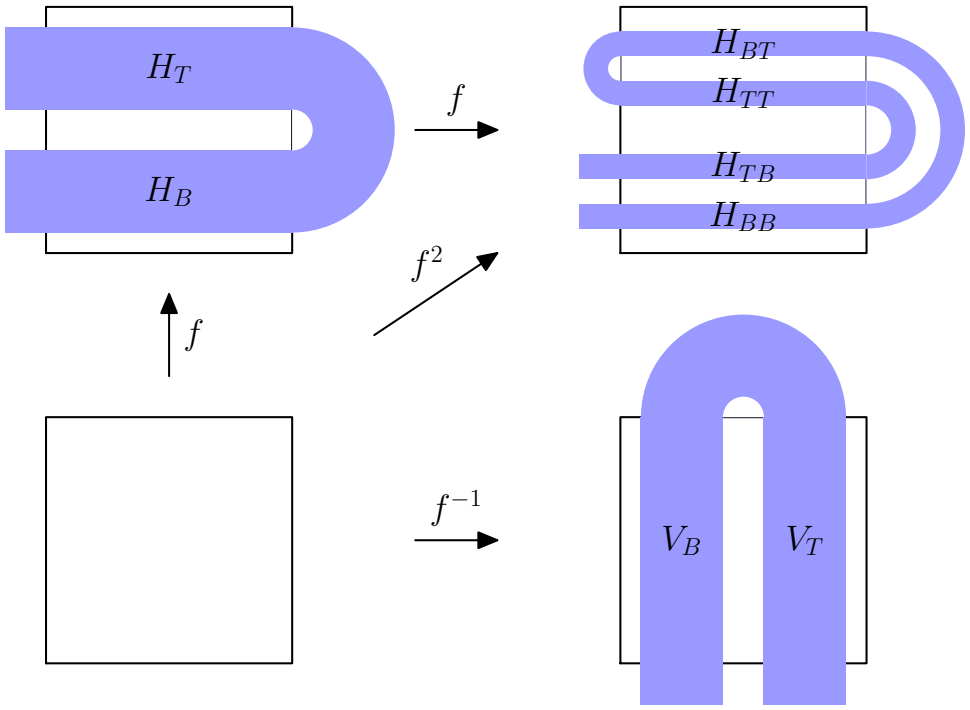


Figure 2.1: A topological picture of the Smale horseshoe map

A horseshoe map f is a diffeomorphism, defined topologically: the map take a square D and stretch it in one direction and squeeze it in another direction to make a long thin strip. Then the strip is placed over the square in the shape of a horseshoe (Figure 2.1).

We study set of the points which remain in the square after some forward and backward

iterations of f. We will call the invariant set

$$\Lambda_N = \{p \mid p \in \bigcap_{n=-N}^{N} f^n(D)\}.$$

When n is sent to ∞, we define $\Lambda = \lim_{n\to\infty} \Lambda_n$.

From Figure 2.1, we see that after one forward iteration of the map, $f(D) \cap D$ consists of two horizontal strips H_T and H_B. If we iterate f once more, all but the four thinner horizontal strips stay inside the square. Observe that two thinner strips are nested in H_T and H_B respectively. We define each thinner nested strip

$$f(H_{s_{-2}}) \cap H_{s_{-1}} = H_{s_{-2}s_{-1}} \text{ where } s_i \in \{T, B\}.$$

Following this way, we label the four horizontal strips $H_{BT}, H_{TT}, H_{TB}, H_{BB}$. If we iterate f forward for n times, then only 2^n horizontal strips remain in D. We inductively define the 2^n horizontal strips,

$$f(H_{s_{-n}\ldots s_{-2}}) \cap H_{s_{-1}} = H_{s_{-n}\ldots s_{-1}} \text{ where } s_i \in \{T, B\}.$$

Now we look at the backward iteration of f. We denote the preimages of the horizontal strips H_T and H_B as $V_T = f^{-1}(H_T)$ and $V_B = f^{-1}(H_B)$. When we iterate f backward for n times, then $f^{-n}(D) \cap D$ is the union of 2^n vertical strips. The nested vertical strips are defined

$$f^{-1}(V_{s_1\ldots s_n}) \cap V_{s_0} = V_{s_0\ldots s_n} \text{ where } s_i \in \{T, B\}.$$

We see that Λ_n is the intersection of 2^n horizontal strips and 2^n vertical strips. When $n \to \infty$, Λ consists of the intersection points of infinite number of horizontal curves and vertical curves, resulting a Cantor set structure. We assign a sequence $s = (\ldots s_{-2}s_{-1}s_0s_1\ldots)$ to each point $p \in \Lambda$ based on to which horizontal curve $H_{\ldots s_{-2}s-1}$ and vertical curve $V_{s_0s_1\ldots}$ the point p belongs. From the construction, this assignment is clearly one-to-one.

Moreover, there is a relation between the dynamics of the points in the invariant Cantor set and a shift map of the sequences. The horseshoe map acting on the invariant set is topologically conjugate to a shift map on the set of infinite sequences of two symbols.

The sequence space provides a convenient setting to examine the periodicity of the orbits. It follows that the horseshoe map contains:

1. Countably many periodic orbits of arbitrarily long period,
2. Uncountably many non-periodic orbits, and
3. A dense orbit.

2.2.3 The Conley-Moser conditions

In general, the image of the "horseshoe map" needs not be in the shape of a horseshoe. The key features of the horseshoe map are the following.

1. The map has stretching and contracting aspects in two different directions.
2. Disjoint regions get mapped to disjoint regions in the domain.

There is a formal criteria called the Conley-Moser conditions, which guarantee the existence of a horseshoe if the conditions are satisfied. To state the conditions, we first need some definitions.

Consider a rectangular domain $D = [X_0, X_1] \times [Y_0, Y_1] \in \mathbb{R}^2$.

Definition. In D, a μ_h-*horizontal curve* is the graph of a function $Y = h(X)$ such that

(1) $Y_0 \leq h(X) \leq Y_1$, and

(2) for every pair $X_i, X_j \in [X_0, X_1]$, we have $|h(X_i) - h(X_j)| \leq \mu_h |X_i - X_j|$ for some $0 \leq \mu_h$.

Definition. In D, a μ_v-*vertical curve* is the graph of a function $X = v(Y)$ such that

(1) $X_0 \leq v(Y) \leq X_1$, and

(2) for every pair $Y_i, Y_j \in [Y_0, Y_1]$, we have $|v(Y_i) - v(Y_j)| \leq \mu_v |Y_i - Y_j|$ for some $0 \leq \mu_v$.

Definition. Given two non-intersecting horizontal curves $h_1(X) < h_2(X)$ in D, we define a μ_h-*horizontal strip* as

$$H = \{(X, Y) : Y \in [h_1(X), h_2(X)] \text{ for } X \in [X_0, X_1]\}.$$

Given two non-intersecting vertical curves $v_1(Y) < v_2(Y)$ in S, we define a μ_v-*vertical strip* as

$$V = \{(X, Y) : Y \in [v_1(Y), v_2(Y)] \text{ for } Y \in [Y_0, Y_1]\}.$$

Definition. The width of horizontal strips is defined as $d(H) = \max |h_2(X) - h_1(X)|$, and the width of vertical strips is defined as $d(V) = \max |v_2(Y) - v_1(Y)|$.

Consider a diffeomorphism $f : D \to \mathbb{R}^2$ and let $S = \{1, 2, \ldots N\}$ be an index set. Let $\bigcup_{s \in S} H_s$ be a set of disjoint μ_h-horizontal strips, and let $\bigcup_{s \in S} V_s$ be a set of disjoint μ_v-vertical strips. The *Conley-Moser conditions* on f are:

CM1 $0 \leq \mu_h \mu_v < 1$

CM2 f maps V_s homeomorphically onto H_s. Also, the horizontal (vertical) boundaries of V_s get mapped to the horizontal (vertical) boundaries of H_s.

CM3 Suppose H is a μ_h-horizontal strip contained in $\bigcup_{s \in S} H_s$, then $f(H) \cap H_s$ is a μ_h-horizontal strip and $d(f(H) \cap H_s) < d(H)$. Similarly, if V is a μ_v-vertical strip contained in $\bigcup_{s \in S} V_i$, then $f^{-1}(V) \cap V_s$ is a μ_v-vertical strip and $d(f^{-1}(V) \cap V_s) < d(V)$.

Theorem 2.4 (Moser)**.** *Suppose f satisfies the Conley-Moser conditions. Then f has an invariant Cantor set Λ on which it is topologically conjugate to a full shift on N symbols, i.e. the*

diagram below commutes, where $\phi : \Lambda \to \Sigma^N$ is a homeomorphism and $\sigma : \Sigma^N \to \Sigma^N$ is the shift map on the space of sequences of N symbols.

$$\begin{array}{ccc} \Lambda & \xrightarrow{\quad f \quad} & \Lambda \\ \Big\downarrow{\scriptstyle \phi} & & \Big\downarrow{\scriptstyle \phi} \\ \Sigma^N & \xrightarrow[\quad \sigma \quad]{} & \Sigma^N \end{array}$$

2.3 Mathematical billiards

A *billiard domain* $\mathcal{Q}$ is a Riemannian manifold with a piecewise smooth boundary $\partial\mathcal{Q}$. A *billiard* is the geodesic flow of a mass point in $\mathcal{Q}$. From the view of classical mechanics, geodesic is a curve satisfying the Hamilton's principle of least action. At the boundary $\partial\mathcal{Q}$, the mass point reflects such that the normal component of the velocity to the boundary reverses its sign and the tangential component of the velocity stays the same.

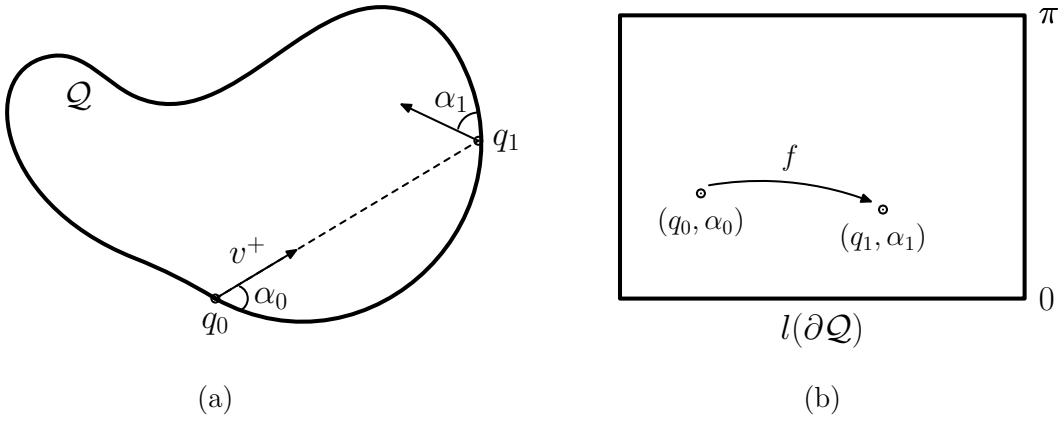


Figure 2.2: (a) A planar billiard and the billiard map. (b) The billiard map on its phase space.

Commonly considered billiard domains are in the plane, since we can explore various mathematical questions even in this simple setting. For planar billiards, a mass point moves in a straight line inside a domain $\mathcal{Q}$ and makes mirror-like reflections at the boundary $\partial\mathcal{Q}$. It is sufficient to study a billiard flow by observing it at the moments of impact. A billiard flow has a natural Poincaré section, which has coordinates (q, α), where q is the arc

length parameter along $\partial\mathcal{Q}$ and α is the angle between the outgoing velocity vector v_+ at the footpoint q and the boundary $\partial\mathcal{Q}$ (Figure 2.2a). The Poincaré section is topologically equivalent to a cylinder $S^1 \times [0, \pi]$. The Poincaré return map, which we call the *billiard map* f, sends a collision (q_0, α_0) to the next collision (q_1, α_1) (Figure 2.2b). If $(q_n, \alpha_n) := f^n(q_0, \alpha_0) = (q_0, \alpha_0)$, we call the set $\{(q_0, \alpha_0), (q_1, \alpha_1), \cdots, (q_n, \alpha_n)\}$ an n-period orbit.

The shape of a billiard domain influences the qualitative behavior of the billiard trajectories in it. We can divide the domains into roughly three different categories. The first type is of smooth convex billiards, called Birkhoff billiards, which are associated to stable dynamics. The common topics in Birkhoff billiards are periodic orbits and the integrability. The second type of billiards are hyperbolic billiards, which often consist of piecewise dispersing boundaries. The most famous hyperbolic billiard is probably the Sinai billiard whose domain is the complement of an array of circles in the plane. As we have seen in the horseshoe map, the hyperbolicity produces chaotic dynamics. The last type of billiards are the ones with polygonal domains. We will not say much about these, except we mention the *unfolding technique* which is at times used to study the billiard trajectories in polygons. Here, we explain the technique using a straight wedge instead.

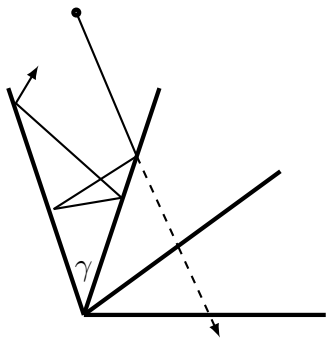


Figure 2.3: An illustration of the unfolding technique in a straight wedge of angle γ.

Suppose a billiard trajectory enters a straight wedge of angle γ and hits one side of the wedge. Instead of reflecting the trajectory at the collision, we reflect the wedge across the side and "unfold" the trajectory into a straight line (Figure 2.3). If we perform this procedure at every collision, then we will get a straight trajectory passing through $\lceil \frac{\pi}{\gamma} \rceil$ wedges. Therefore, the maximal number of collisions inside the wedge of angle γ is $\lceil \frac{\pi}{\gamma} \rceil$.

CHAPTER 3

The dumbbell system

In the two-dimensional space with coordinates (X, Y), let us consider a dumbbell system, which consists of two point masses m_{L}, m_{R}, connected by weightless rigid rod of length l. The coordinates of m_{L}, m_{R}, and the center of mass of the dumbbell are denoted by $(X_{\mathrm{L}}, Y_{\mathrm{L}}), (X_{\mathrm{R}}, Y_{\mathrm{R}})$, and (X, Y), respectively. Let θ be the angular position of the vector $\overrightarrow{X_{\mathrm{L}}X_{\mathrm{R}}}$ measured counterclockwise from the positive X-direction (Figure 3.1a). With the total mass of the dumbbell $m = m_{\mathrm{L}} + m_{\mathrm{R}}$, we define $\beta_{\mathrm{L}} = \frac{m_{\mathrm{L}}}{m}$ and $\beta_{\mathrm{R}} = \frac{m_{\mathrm{R}}}{m}$. Then the distance from the center of mass to m_{L} and to m_{R} are given by $l\beta_{\mathrm{R}}$ and $l\beta_{\mathrm{L}}$.

The dumbbell moves freely in the half space $Y > 0$ until it reflects elastically on the floor $Y = 0$. The velocity of the center of mass in the X-direction is constant since there is no force acting on the system in that direction. Thus, we may assume that the center of mass does not move in the X-direction. With this reduction, the dumbbell configuration space is a two-dimensional cylinder with the natural choice of coordinates (θ, Y).

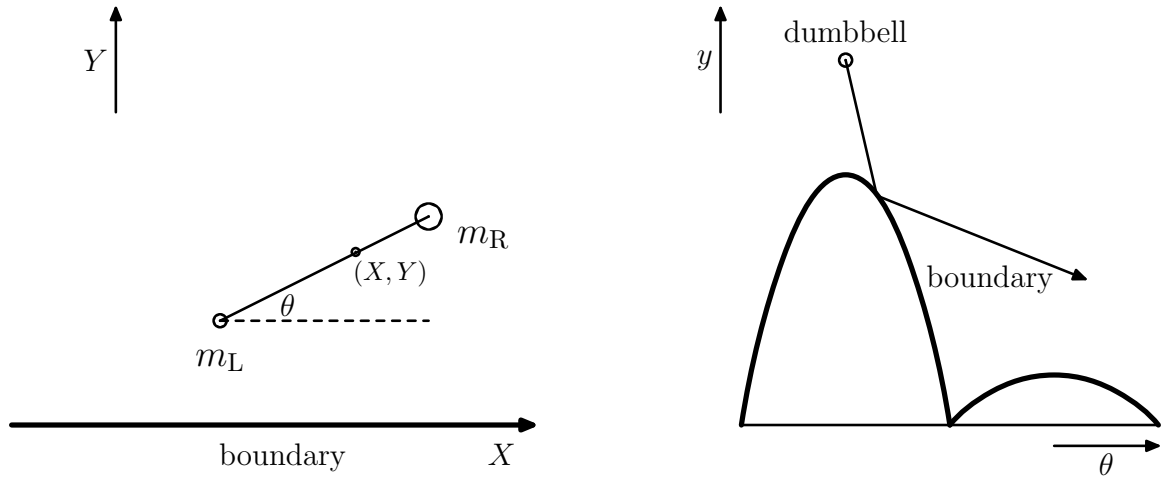
On the θY plane, the dumbbell is a mass point in a transformed domain. Note that on the XY plane, the dumbbell hits the floor when Y_{L} or Y_{R} becomes zero. Therefore, using the relations

$$Y_{\mathrm{L}} = Y - l\beta_{\mathrm{R}} \sin\theta \geq 0 \qquad\qquad Y_{\mathrm{R}} = Y + l\beta_{\mathrm{L}} \sin\theta \geq 0,$$

we find the domain on the θY plane

$$\{(\theta, Y) : Y \geq \max\{l\beta_{\mathrm{R}} \sin\theta, -l\beta_{\mathrm{L}} \sin\theta\} \text{ for } \theta \in [0, 2\pi)\}.$$

Now we examine the motion of the dumbbell. We assume no gravity in the system. The



(a) (b)

Figure 3.1: (a) The dumbbell model with the flat boundary. (b) The billiard system for the dumbbell.

moment of inertia of the dumbbell is given by $I = m_{\mathrm{L}}(l\beta_{\mathrm{R}})^2 + m_{\mathrm{R}}(l\beta_{\mathrm{L}})^2 = ml^2\beta_{\mathrm{L}}\beta_{\mathrm{R}}$. It is convenient to rescale $Y = \sqrt{\frac{I}{m}}y$ since in this way the Lagrangian is equal to the that of the free particle

$$L = \frac{I\dot{\theta}^2}{2} + \frac{m\dot{Y}^2}{2} = \frac{I\dot{\theta}^2}{2} + \frac{I\dot{y}^2}{2}. \tag{3.1}$$

By Hamilton's principle of least action, true orbits extremize $\int_{t_0}^{t_1} L(\dot{y}, \dot{\theta})\mathrm{d}t$. We see that the billiard trajectories on the θy plane are straight lines between two collisions.

We also rescale the billiard domain accordingly and get

$$\mathcal{Q} = \left\{(\theta, y) : y \geq \max\left\{\sqrt{\frac{\beta_{\mathrm{R}}}{\beta_{\mathrm{L}}}}\sin\theta, -\sqrt{\frac{\beta_{\mathrm{L}}}{\beta_{\mathrm{R}}}}\sin\theta\right\} \text{ for } \theta \in [0, 2\pi)\right\}. \tag{3.2}$$

The boundary $\partial\mathcal{Q}$ has non-smooth corners at $\theta = 0, \pi$. This is the case when the dumbbell's two masses hit the floor at the same time. We will not consider this degenerate case in our paper.

At the boundary $\partial\mathcal{Q}$, there will be a mirror-like reflection. We use the subscript _ to

denote the values right before a collision and $_+$ to denote the values right after a collision. Given the incoming velocity vector $\mathbf{v}_-$ and normal vector $\mathbf{n}$ to the boundary $\partial\mathcal{Q}$, the outgoing velocity vector $\mathbf{v}_+$ can be found using

$$\mathbf{v}_+ = -2\frac{\mathbf{v}_- \cdot \mathbf{n}}{\mathbf{n} \cdot \mathbf{n}}\mathbf{n} + \mathbf{v}_-. \tag{3.3}$$

In the case when m_{L} hits the boundary, we have

$$\mathbf{v}_- = [\dot{\theta}_-, \dot{y}_-] \qquad \mathbf{n} = \left[-\sqrt{\frac{\beta_{\mathrm{R}}}{\beta_{\mathrm{L}}}}\cos\theta, 1 \right].$$

Then, using (3.3), we compute $\mathbf{v}_+ = [\dot{\theta}_+, \dot{y}_+]$. Changing back to the original coordinates (θ, Y), we express the translational and the angular velocities of the dumbbell after the collision in terms of the velocities before the collision.

$$\begin{pmatrix} \dot{\theta}_+ \\ \dot{Y}_+ \end{pmatrix} = \begin{pmatrix} \dot{\theta}_- \left(1 - \dfrac{2\beta_{\mathrm{R}}\cos^2\theta}{\beta_{\mathrm{L}} + \beta_{\mathrm{R}}\cos^2\theta} \right) + \dot{Y}_- \left(\dfrac{2/l\cos\theta}{\beta_{\mathrm{L}} + \beta_{\mathrm{R}}\cos^2\theta} \right) \\ \dot{\theta}_- \left(\dfrac{2l\beta_{\mathrm{L}}\beta_{\mathrm{R}}\cos\theta}{\beta_{\mathrm{L}} + \beta_{\mathrm{R}}\cos^2\theta} \right) + \dot{Y}_- \left(1 - \dfrac{2\beta_{\mathrm{L}}}{\beta_{\mathrm{L}} + \beta_{\mathrm{R}}\cos^2\theta} \right) \end{pmatrix}. \tag{3.4}$$

Remark 3.1. The post-collision velocities for the other case, when m_{R} hits the boundary, can be obtained in a similar manner: we switch β_{L} and β_{R}, replace $\cos\theta$ and $\sin\theta$ with $-\cos\theta$ and $-\sin\theta$, and replace Y_{L} with Y_{R}.

CHAPTER 4

Estimate on the number of collisions of the dumbbell

Consider the dumbbell with the arbitrary masses m_{L} and m_{R} moving in the XY plane where no gravity force is present. We estimate the maximal number of collisions of the dumbbell with the floor as a function of the mass ratios. As we have seen in Chapter 3, on the θy plane, the dumbbell reduces to a mass point that has unit velocity in a billiard domain on a cylinder and makes elastic reflections at the boundary.

First, we discuss the properties of the billiard boundary for the dumbbell system on the θy plane. Recall from (3.2) that the boundary is given by

$$\mathcal{Q} = \left\{ (\theta, y) : y \geq \max \left\{ \sqrt{\frac{\beta_{\mathrm{R}}}{\beta_{\mathrm{L}}}} \sin\theta, -\sqrt{\frac{\beta_{\mathrm{L}}}{\beta_{\mathrm{R}}}} \sin\theta \right\} \text{ for } \theta \in [0, 2\pi) \right\}. \tag{4.1}$$

When $m_{\mathrm{L}} = m_{\mathrm{R}}$, we have $\beta_{\mathrm{L}} = \beta_{\mathrm{R}} = \frac{1}{2}$ and the boundary is $y = |\sin\theta|$. The angle between the two sine waves is $\frac{\pi}{2}$. When $m_{\mathrm{L}} \neq m_{\mathrm{R}}$, it follows from (4.1) that the boundary consists of two sine waves with different heights. We will assume $m_{\mathrm{L}} < m_{\mathrm{R}}$, since the case $m_{\mathrm{R}} > m_{\mathrm{L}}$ is symmetric. To visualize the billiard boundary more conveniently, we measure θ in the range $[\pi/2, 5\pi/2)$ instead of $[0, 2\pi)$ (Figure 4.1). It is easy to see that generically in the limit $\frac{m_{\mathrm{L}}}{m_{\mathrm{R}}} \to 0$ most of repeated collisions will occur between the two tall parts of the boundary.

Let us consider the straight wedge formed by the tangent lines to $y = \sqrt{\frac{\beta_{\mathrm{R}}}{\beta_{\mathrm{L}}}} \sin\theta$ at $\theta = \pi$ and 2π. We call these tangent lines ℓ_π and $\ell_{2\pi}$ respectively. We denote the angle of the straight wedge by γ. We define the hybrid wedge as the union of the sine waves when $y > 0$ and the tangent lines ℓ_π and $\ell_{2\pi}$ when $y \leq 0$. See Figure 4.1 for the construction of the wedges.

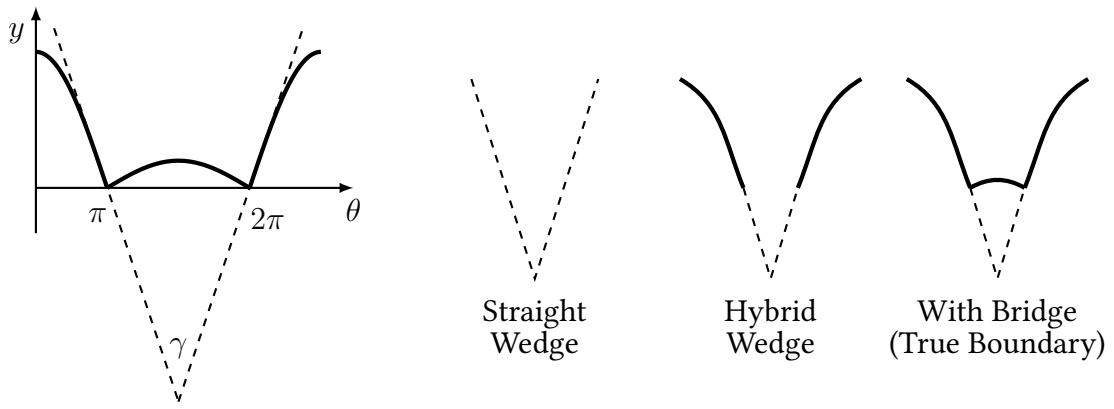


Figure 4.1: The construction of the three different types of boundaries.

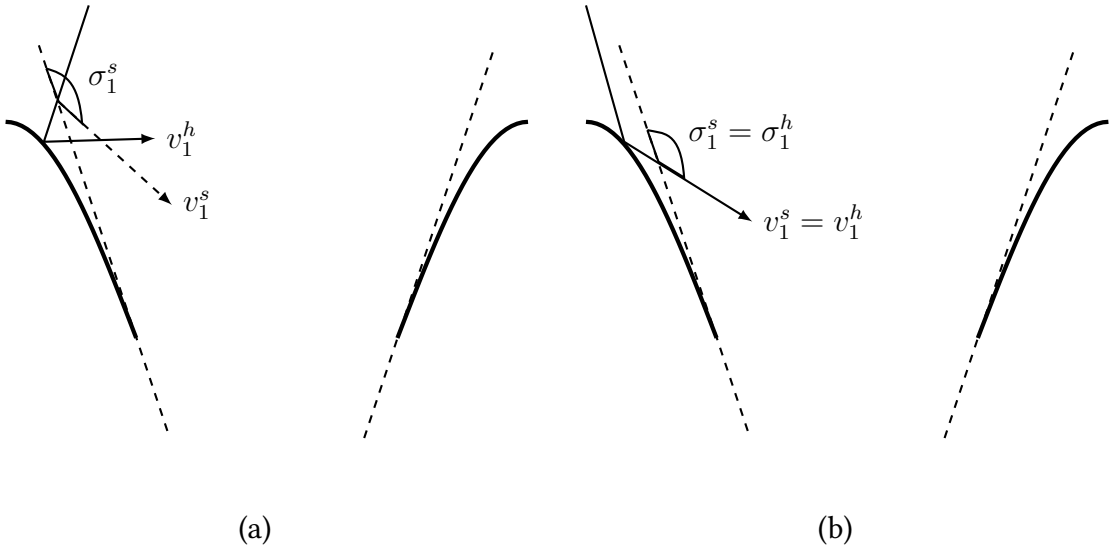


Figure 4.2: Two different base cases for Lemma 4.1.

We introduce some notations. We denote a billiard trajectory interacting with the hybrid wedge by v^h. Similarly, a trajectory bouncing inside the straight wedge is called v^s. When v^h or v^s is written with the subscript i, it denotes the segment of the corresponding trajectory between the i-th collision and the $i+1$-st collision. Let σ_i^h be the angle from the straight wedge to v_i^h, and σ_i^s denote the angle from the straight wedge to v_i^s after the i-th collision. Define ρ_i^h as the angle deviation of the hybrid wedge from the straight wedge at the i-th collision of v_i^h.

The trajectory v^h or v^s will terminate when the sequence of angles $\{\sigma_i^h\}$ or $\{\sigma_i^s\}$ termi-

nates (due to the absence of the next bounce), or when there will be no more intersections with the straight wedge. This will happen when the angle of intersection, σ_i^h or σ_i^s, with the straight wedge is less than or equal to γ.

Lemma 4.1. *Consider the hybrid wedge and the straight wedge described above. The sequence of angles $\sigma_i^h, 1 \leq i$ will terminate after or at the same index as the sequence of angles $\sigma_i^s, 1 \leq i$.*

Proof. Suppose that the initial segment $v_0^s = v_0^h$ of the dumbbell trajectory crosses the straight wedge before it hits the hybrid wedge, as shown in Figure 4.2a. If the incoming angle is α, then

$$\sigma_1^h = \pi - \alpha - 2\rho_1^h < \pi - \alpha = \sigma_1^s.$$

When the initial segment hits the hybrid wedge before crossing the straight wedge, as shown in Figure 4.2b, then set $\sigma_1^h = \sigma_1^s$. Now we can proceed by induction if $\sigma_i^h, \sigma_i^s > \gamma$ and if the sequence σ_i^h has not terminated.

$$\sigma_{i+1}^h = \sigma_i^h - \gamma - 2\rho_{i+1}^h$$
$$\sigma_{i+1}^s = \sigma_i^s - \gamma$$

which implies that $\sigma_{i+1}^h \leq \sigma_{i+1}^s$.
Since $\sigma_i^h \leq \sigma_i^s$, then v^h will terminate at the same time or before v^s. □

Define the bridge as the smaller sine wave created by $y = -\sqrt{\frac{\beta_{\mathrm{L}}}{\beta_{\mathrm{R}}}} \sin\theta$ for $\theta \in [\pi, 2\pi)$. The union of the bridge with the hybrid wedge is the true boundary defined by the dumbbell dynamics (Figure 4.1).

Lemma 4.2. *The presence of the bridge in the hybrid wedge will increase the number of collisions of the dumbbell by at most one from the number of collisions of the dumbbell to the hybrid wedge.*

Proof. Consider the true trajectory v that "sees" the bridge. Recall the definitions of σ_i^h

and ρ_i^h. Similarly, we let σ_i be the angle to v_i and let ρ_i be the angle deviation of the true boundary from the straight wedge at the i-th collision of v_i.

Before v intersects the bridge, by Lemma 4.1 we have

$$v_i = v_i^h, \quad \sigma_i = \sigma_i^h, \quad \text{and} \quad \sigma_i^h = \sigma_{i-1}^h - \gamma - 2\rho_i^h.$$

Now define τ to be the signed angle measured from the horizontal line to the tangent line of the bridge curve where v hits the bridge (Figure 4.3). Note that τ takes a positive value if v hits the left half of the bridge, and τ takes a negative value if v hits the right half of the bridge. We express σ_{i+1} after the bounce from the bridge in terms of σ_i. By this convention, the bounce from the bridge does not increase the index count. We will adjust the index count by adding $+1$ at the end.

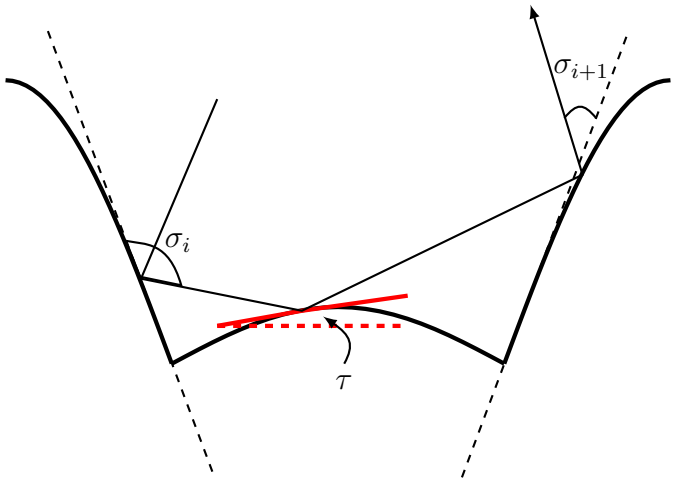


Figure 4.3: The real billiard trajectory v that sees the bridge.

Then, we have

$$\begin{aligned} \sigma_{i+1} &= \pi - \sigma_i - 2\tau - 2\rho_{i+1} \\ \sigma_{i+1}^h &= \sigma_i^h - \gamma - 2\rho_{i+1}^h \end{aligned} \tag{4.2}$$

We may assume that v hits the bridge with non-positive velocity in y. If the dumbbell hits the bridge with positive velocity in y, it will continue to move in the positive y direction after reflection from the bridge. Then, we consider the reverse trajectory to bound the

number of collisions. This allows us to restrict σ_i. Moreover, v_i naturally hits the upper part of hybrid wedge than v_i^h. We also assume that v hits the left half of the bridge. Otherwise, we can reflect the orbit around the vertical line passing through the middle point of the bridge.

Utilizing the above arguments, we have the inequalities

$$\frac{\pi+\gamma}{2} < \sigma_i, \quad 0 < \tau < \frac{\gamma}{2}, \quad \text{and} \quad \rho_{i+1}^h < \rho_{i+1}. \tag{4.3}$$

It is straightforward to verify that (4.2) and (4.3) imply $\sigma_{i+1} \leq \sigma_{i+1}^h$.

From the $i+2$-nd bounce, if σ_i has not terminated, we can apply the induction argument similar to the proof of Lemma 4.1. We have the base case

$$\sigma_{i+1} \leq \sigma_{i+1}^h, \qquad \text{and} \qquad \rho_{i+1} \geq \rho_{i+1}^h.$$

Note that ρ's indicate the relative position of a collision point in the hybrid wedge. That is, if $\rho_{i+1} \geq \rho_{i+1}^h$, then the starting point of v_{i+1} is located at or above that of v_{i+1}^h. Since $\sigma_{i+1} \leq \sigma_{i+1}^h$ and v_{i+1} starts above v_{i+1}^h, we know v_{i+2} will start on the hybrid wedge higher than v_{i+2}^h. This implies $\rho_{i+2} \geq \rho_{i+2}^h$. Then using the recursive relationship,

$$\begin{aligned}\sigma_{i+2} &= \sigma_{i+1} - \gamma - 2\rho_{i+2} \\ \sigma_{i+2}^h &= \sigma_{i+1}^h - \gamma - 2\rho_{i+2}^h,\end{aligned}$$

we obtain $\theta_{i+2} \leq \theta_{i+2}^h$. By induction $\theta_i \leq \theta_i^h$ for all i. Taking into account the bounce on the bridge, we conclude that the number of bounces of v will increase at most by one relative to that of v^h. □

Remark 4.3. In most cases, the number of bounces of v will be less than the number of bounces of v^h (Figure 4.4).

Now we are ready to prove the main theorem of the chapter.

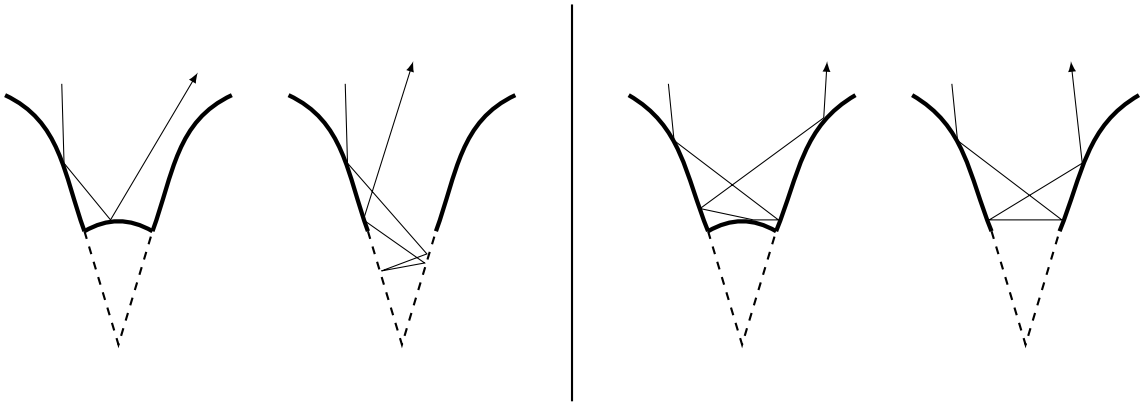


Figure 4.4: (a) The bridge prevents the trajectory to travel deeper. (b) The bridge increases the number of collisions by 1.

Theorem 4.4. *The number of collisions of the dumbbell is bounded above by* $N_\gamma = \left\lceil \frac{\pi}{\gamma} \right\rceil + 1$, *where* $\gamma = \pi - 2\arctan\sqrt{\frac{m_{\mathrm{R}}}{m_{\mathrm{L}}}}$.

Proof. When $m_{\mathrm{L}} = m_{\mathrm{R}}$, the boundary consists of two equal sine waves which form hybrid wedges of angle $\frac{\pi}{2}$ at $\theta = \pi, 2\pi$. Using Lemma 4.1, we conclude that the upper bound for the number of collisions is $\lceil \frac{\pi}{\gamma} \rceil = 2$, which is less than $N_\gamma = 3$.

When $m_{\mathrm{L}} < m_{\mathrm{R}}$, we consider the true boundary which consists of a hybrid wedge with the bridge. By Lemma 4.1 and Lemma 4.2, the the maximal number of collisions to the true boundary is bounded above by $N_\gamma = \lceil \frac{\pi}{\gamma} \rceil + 1$. Since $\gamma = \pi - 2\arctan\sqrt{\frac{\beta_{\mathrm{R}}}{\beta_{\mathrm{L}}}} = \pi - 2\arctan\sqrt{\frac{m_{\mathrm{R}}}{m_{\mathrm{L}}}}$, this completes the proof. □

CHAPTER 5

Adiabatic invariance in the dumbbell

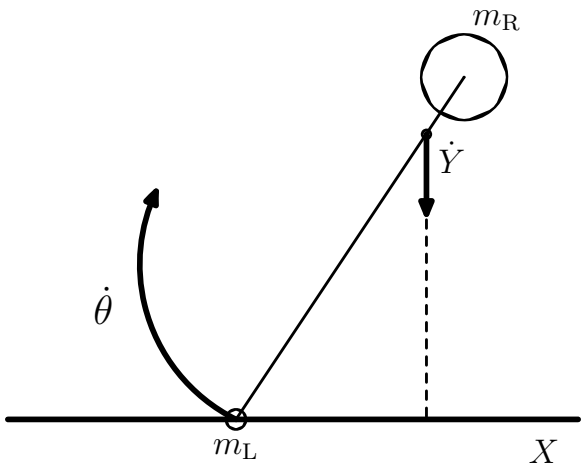


Figure 5.1: The light mass bounces many times off the floor while the large mass slowly approaches the floor.

We consider the case when the dumbbell has $m_{\mathrm{L}} \ll m_{\mathrm{R}}$ and m_{L} rotates around m_{R} with a high angular velocity $\dot{\theta}$. For easy computations, we set the length of the rod $l = 1$. We further assume that the center of mass has a slow downward velocity compared to $\dot{\theta}$. Since multiplying the velocities $(\dot{\theta}, \dot{Y})$ by a constant does not change the orbit, we normalize $\dot{\theta}$ to be of order 1, then $\dot{Y}$ is small. slowly approaching the floor, rotating with angular velocity of order 1.

At some moment the small mass m_{L} will hit the floor. If the angle $\theta = \frac{\pi}{2}$ (or sufficiently close to it), then the dumbbell will bounce away without experiencing any more collisions. This situation is rather exceptional.

A simple calculation shows that $|\theta - \frac{3\pi}{2}|$ will be generically of order $\sqrt{|\dot{Y}|}$ for our limit $\dot{Y} \to 0$. In this chapter we assume this favorable scenario. For the corresponding set of

initial conditions, we obtain an adiabatic invariant. We start by deriving approximate map between two consecutive bounces. To avoid cumbersome use of implicit constants, we use some standard notation when dealing with asymptotic expansions:

Notation:

$f \lesssim g \Leftrightarrow f = O(g) \Leftrightarrow f \leq Cg$ for some $C > 0$

$f \gtrsim g \Leftrightarrow g = O(f)$

$f \sim g \Leftrightarrow f \lesssim g$ and $f \gtrsim g$.

Lemma 5.1. *Let $\beta_{\mathrm{L}} = \epsilon \ll 1$, $\dot{\theta}_- \neq 0$ and assume m_{L} bounces off the floor and hits the floor next before m_{R} does. Then there exist $\delta = \delta(\epsilon)$ such that $\delta/\epsilon \to \infty$ and $\delta \to 0$ as $\epsilon \to 0$, and if $-\delta < \dot{Y}_- < 0$ and $|\phi - \frac{3\pi}{2}| \gtrsim \sqrt{\delta}$, the collision map is given by*

$$\dot{\theta}_+ = -\dot{\theta}_- - \frac{2}{\sqrt{1 - Y_-^2}}\dot{Y}_- + O\left(\frac{\epsilon}{\delta}\right) \tag{5.1}$$

$$Y_+ = Y_- - \frac{2\pi - 2\arccos Y_-}{\dot{\theta}_-}\dot{Y}_- + O\left(\delta^{3/2}\right) + O\left(\frac{\epsilon}{\sqrt{\delta}}\right). \tag{5.2}$$

Proof. We prove (5.1) in two steps. We first show that

$$\dot{\theta}_+ = -\dot{\theta}_- - \frac{2}{\cos\phi}\dot{Y}_- + O\left(\frac{\epsilon}{\delta}\right)$$

using the expression for $\dot{\theta}_+$ in (3.4). We have,

$$\begin{aligned}
&\dot{\theta}_+ + \left(\dot{\theta}_- + \frac{2}{\cos\theta}\dot{Y}_-\right) \\
&= \left(1 - \frac{2(1-\beta_{\mathrm{L}})\cos^2\theta}{\beta_{\mathrm{L}} + (1-\beta_{\mathrm{L}})\cos^2\theta}\right)\dot{\theta}_- + \left(\frac{2\cos\theta}{\beta_{\mathrm{L}} + (1-\beta_{\mathrm{L}})\cos^2\theta}\right)\dot{Y}_- + \left(\dot{\theta}_- + \frac{2}{\cos\theta}\dot{Y}_-\right) \\
&= \frac{(\beta_{\mathrm{L}} - (1-\beta_{\mathrm{L}})\cos^2\theta)\dot{\theta}_- - (2\cos\theta)\dot{Y}_-}{\beta_{\mathrm{L}}\sin^2\theta + \cos^2\theta} + \left(\dot{\theta}_- + \frac{2}{\cos\theta}\dot{Y}_-\right) \\
&= \beta_{\mathrm{L}}\left(\frac{\dot{\theta}_- + 1 + 2\dot{Y}_-(\frac{\sin^2\theta}{\cos\theta})}{\beta_{\mathrm{L}}\sin^2\theta + \cos^2\theta}\right).
\end{aligned}$$

For sufficiently small δ, $\left|\theta - \frac{3\pi}{2}\right| \gtrsim \sqrt{\delta}$ implies $\cos\theta \gtrsim \sqrt{\delta}$. It follows that

$$\left|\dot{\theta}_+ + \left(\dot{\theta}_- + \frac{2}{\cos\theta}\dot{Y}_-\right)\right| \lesssim \epsilon \left|\frac{1 + \frac{2\delta}{\sqrt{\delta}}}{\delta}\right| = O\left(\frac{\epsilon}{\delta}\right).$$

Observe from the Figure 5.1 that $\theta = \dfrac{3\pi}{2} - \arccos\left(\dfrac{Y_-}{1-\beta_{\mathrm{L}}}\right)$. Thus,

$$\begin{aligned}\dot{\theta}_+ + \left(\dot{\theta}_- + \frac{2}{\cos\theta}\dot{Y}_-\right) &= \dot{\theta}_+ + \dot{\theta}_- + \frac{2}{\cos\left(\dfrac{3\pi}{2} - \arccos\left(\dfrac{Y_-}{1-\beta_{\mathrm{L}}}\right)\right)}\dot{Y}_- \\ &= \dot{\theta}_+ + \dot{\theta}_- + \frac{2}{\sqrt{1 - \left(\dfrac{Y_-}{1-\beta_{\mathrm{L}}}\right)^2}}\dot{Y}_- = \dot{\theta}_+ + \dot{\theta}_- + \frac{2}{\sqrt{1 - Y_-^2}}\dot{Y}_- + R_1,\end{aligned}$$

where

$$|R_1| \le \beta_{\mathrm{L}} \left|\frac{2\dot{Y}_- Y_-}{\sqrt{((1-\beta_{\mathrm{L}})^2 - Y_-^2)^3}}\right|.$$

Using that $\sqrt{(1-\beta_{\mathrm{L}})^2 - Y_-^2} = (1-\beta_{\mathrm{L}})\cos\theta$, we obtain

$$|R_1| \lesssim \epsilon \left|\frac{2\delta}{((1-\beta_{\mathrm{L}})\sqrt{\delta})^3}\right|.$$

Combining the results, we have

$$\begin{aligned}\left|\dot{\theta}_+ + \dot{\theta}_- + \frac{2}{\sqrt{1 - Y_-^2}}\dot{Y}_-\right| &= \left|\dot{\theta}_+ + \left(\dot{\theta}_- + \frac{2}{\cos\theta}\dot{Y}_-\right)\right| + |R_1| \\ &\lesssim \epsilon \left(\left|\frac{1 + \frac{2\delta}{\sqrt{\delta}}}{\delta}\right| + \left|\frac{2\delta}{((1-\beta_{\mathrm{L}}))^3\delta^{3/2}}\right|\right) = O\left(\frac{\epsilon}{\delta}\right).\end{aligned}$$

This completes the proof for (5.1).

Let t be the time between the two consecutive collisions of m_{L}. Then $Y_+ = Y_- - \dot{Y}_- t$.

The angular distance that m_{L} traveled is given by

$$\begin{aligned}\psi &= 2\pi - \arccos\left(\frac{Y_-}{1-\beta_{\mathrm{L}}}\right) - \arccos(Y_+) \\ &= 2\pi - \arccos\left(\frac{Y_-}{1-\beta_{\mathrm{L}}}\right) - \arccos\left(\frac{Y_- - \dot{Y}_- t}{1-\beta_{\mathrm{L}}}\right) \\ &= 2\pi - 2\arccos\left(\frac{Y_-}{1-\beta_{\mathrm{L}}}\right) + R_2 = 2\pi - 2\arccos Y_- + R_3 + R_2,\end{aligned}$$

where R_2 and R_3 are the error estimates for the Taylor series expansion and are given explicitly by

$$\begin{aligned}|R_2| &\le \left|\frac{\dot{Y}_- t}{\sqrt{(1-\beta_{\mathrm{L}})^2 - Y_-^2}}\right| \\ |R_3| &\le \left|\frac{\beta_{\mathrm{L}} Y_-}{(1-\beta_{\mathrm{L}})\sqrt{(1-\beta_{\mathrm{L}})^2 - Y_-^2}}\right|.\end{aligned}$$

Therefore, we have

$$\begin{aligned}Y_+ &= Y_- - \dot{Y}_- t = Y_- - \dot{Y}_-\left(\frac{\psi}{\dot{\theta}_-}\right) \\ &= Y_- - \frac{\dot{Y}_-}{\dot{\theta}_-}\left(2\pi - 2\arccos Y_- + R_2 + R_3\right) \\ &= Y_- - \frac{2\pi - 2\arccos Y_-}{\dot{\theta}_-}\dot{Y}_- + \frac{\dot{Y}_-}{\dot{\theta}_-}(R_2 + R_3).\end{aligned}$$

Since m_{L} can travel at most 2π between two collisions, t is bounded by $|t| < 2\pi/\dot{\theta}$. Also note that R_2 and R_3 contain the factor $\dot{Y}_-$ and β_{L} respectively. We finish the proof for (5.2)

by computing,

$$\begin{aligned}\left|Y_+ - Y_- + \frac{2\pi - 2\arccos Y_-}{\dot\theta_-}\dot Y_-\right| &\lesssim \left|\frac{\dot Y_-}{\dot\theta_-}(R_2+R_3)\right| \\ &\lesssim \dot Y_-^2\left|\frac{2\pi Y_-}{\dot\theta_-\sqrt{(1-\beta_{\mathrm{L}})^2 - Y_-^2}}\right| + \beta_{\mathrm{L}}\left|\frac{2\pi Y_-}{\dot\theta_-(1-\beta_{\mathrm{L}})\sqrt{(1-\beta_{\mathrm{L}})^2 - Y_-^2}}\right| \\ &\lesssim \delta^2\left|\frac{2\pi}{\sqrt{\delta}}\right| + \epsilon\left|\frac{2\pi}{((1-\beta_{\mathrm{L}}))^2\sqrt{\delta}}\right| = O\left(\delta^{3/2}\right) + O\left(\frac{\epsilon}{\sqrt{\delta}}\right).\end{aligned}$$

□

Corollary 5.2. *Under the same assumptions as in Lemma 5.1 with the exception* $\left|\phi - \frac{3\pi}{2}\right| \gtrsim \delta^k$ *for* $0 \le k \le 1$ *and* $\epsilon \ll \delta^{2k}$*, the variables after the collision are given by the similar equations to (5.1) and (5.2) but with different error terms.*

$$\dot\theta_+ = -\dot\theta_- - \frac{2}{\sqrt{1-Y_-^2}}\dot Y_- + O\left(\frac{\epsilon}{\delta^{2k}}\right) \tag{5.3}$$

$$Y_+ = Y_- - \frac{2\pi - 2\arccos Y_-}{\dot\theta_-}\dot Y_- + O\left(\frac{\delta^2}{\delta^k}\right) + O\left(\frac{\epsilon}{\delta^k}\right). \tag{5.4}$$

Proof. When computing the error terms, use $\cos\theta \gtrsim \delta^k$. □

Now, we can state the adiabatic invariance theorem for the special case when the light mass hits the floor and the dumbbell is far away from the vertical position: $\theta = \frac{3\pi}{2}$.

Theorem 5.3. *Suppose right before the collision* $\dot\theta_- \neq 0$ *and* $\theta - \frac{3\pi}{2} \neq 0$*. Then there is* $\delta > 0$ *such that if* $0 < \beta_{\mathrm{L}} = \epsilon = \delta^2$ *and* $-\delta < \dot Y_0 < 0$*, then there exists an adiabatic invariant of the dumbbell system, given by* $I = |\dot\theta| f(Y)$*, where* $f(Y) = \pi - \arccos Y$*. In other words,* $|\dot\theta_n| f(Y_n) - |\dot\theta_0| f(Y_0) = O(\delta)$ *after* $N = O(\delta^{-1})$ *collisions.*

Proof. We prove this by finding $f(Y)$ that satisfies

$$|\dot\theta_+| f(Y_+) - |\dot\theta_-| f(Y_-) = O(\delta^2).$$

When $\epsilon = \delta^2$ and δ is sufficiently small, it follows from (5.4) that,

$$f(Y_+) = f(Y_-) - \left(\frac{2\pi - 2\arccos Y_-}{\dot{\theta}_-} \dot{Y}_- + O(\delta^2) \right) f'(Y_-).$$

Then, we have

$$|\dot{\theta}_+| f(Y_+) = \left| \dot{\theta}_- + \frac{2\dot{Y}_-}{\sqrt{1 - Y_-^2}} + O(\delta^2) \right| \left(f(Y_-) - \left(\frac{2\pi - 2\arccos Y_-}{\dot{\theta}_-} \dot{Y}_- + O(\delta^2) \right) f'(Y_-) \right)$$

$$= |\dot{\theta}_-| f(Y_-) - (2\pi - 2\arccos Y_-)\, \dot{Y}_- f'(Y_-) + \left| \frac{2}{\sqrt{1 - Y_-^2}} \dot{Y}_- \right| f(Y_-) + O(\delta^2).$$

Therefore, $f(Y)$ satisfies $|\dot{\theta}_+| f(Y_+) - |\dot{\theta}_-| f(Y_-) = O(\delta^2)$ provided

$$-(2\pi - 2\arccos Y_-)\dot{Y}_- f'(Y_-) + \frac{2}{\sqrt{1 - Y_-^2}} \dot{Y}_- f(Y_-) = 0.$$

The solution of the above equation is given by $f(Y_-) = \pi - \arccos Y_-$.

Let $N = O(\delta^{-1})$ and let $\dot{\theta}_N$ and Y_N be the angular velocity and the distance after N^{th} collision. Then, we have

$$|\dot{\theta}_N| f(Y_N) - |\dot{\theta}_0| f(Y_0) = \sum_{k=1}^{N} \left(|\dot{\theta}_k| f(Y_k) - |\dot{\theta}_{k-1}| f(Y_{k-1}) \right) \lesssim N \cdot \delta^2 \lesssim \delta.$$

□

Remark 5.4. Adiabatic invariant has a natural geometric meaning: angular velocity times the distance traveled by the light mass between two consecutive collisions.

Now, we state the theorem for a realistic scenario when a rapidly rotating dumbbell scatters off the floor.

Theorem 5.5. *Let the dumbbell approach the floor from infinity with $\dot{\theta}_- \neq 0$. There exists $\delta > 0$ such that if $0 < \beta_{\mathrm{L}} = \epsilon = \delta^2$, $-\delta < \dot{Y}_- < 0$, $|\theta_0 - \frac{3\pi}{2}| \sim \sqrt{\delta}$ then, after $N = O(\delta^{-1})$*

bounces the dumbbell will leave the floor after the final bounce by m_{L} *with* $I_N = I_0 + O(\sqrt{\delta})$. *The adiabatic invariant is defined as above* $I = |\dot{\theta}| f(Y)$.

Remark 5.6. The condition on the angle $|\theta_0 - \frac{3\pi}{2}| \sim \sqrt{\delta}$ comes naturally from the following argument. If $\dot{Y} = -\delta$, the dumbbell approaching from infinity will naturally hit the floor when $y \gtrsim 1 - \beta_{\mathrm{L}} - \delta$. Since δ is small, this implies $|\theta_0 - \frac{3\pi}{2}| \lesssim \sqrt{\delta}$. If θ_0 happens to be too close to $\frac{3\pi}{2}$, then there is no hope to obtain adiabatic invariant and we exclude such set of initial conditions. In the limit $\delta \to 0$ the relative measure of the set where $|\theta_0 - \frac{3\pi}{2}| = o(\sqrt{\delta})$ tends to zero.

Proof. We will split the iterations (bounces) into two parts: before the n^{th} iteration and after it, where $n = \lfloor \frac{\mu}{\sqrt{\delta}} \rfloor$ and μ is sufficiently small (to be defined later). We claim that after n bounces, $|\theta_n - \frac{3\pi}{2}| \gtrsim \sqrt[4]{\delta}$. To prove this claim, we use energy conservation of the dumbbell system (3.1), and (5.3). We have

$$\epsilon(1-\epsilon)\left(\dot{\theta}_- + \frac{2}{\sqrt{1-Y_-^2}}\dot{Y}_- + O\left(\frac{\epsilon}{\delta}\right)\right)^2 + \dot{Y}_+^2 = \epsilon(1-\epsilon)\dot{\theta}_-^2 + \dot{Y}_-^2. \tag{5.5}$$

Next,

$$\begin{aligned}
|\dot{Y}_+^2 - \dot{Y}_-^2| &= \left|\epsilon(1-\epsilon)\dot{\theta}_-^2 - \epsilon(1-\epsilon)\left(\dot{\theta}_- + \frac{2}{\sqrt{1-Y_-^2}}\dot{Y}_- + O\left(\frac{\epsilon}{\delta}\right)\right)^2\right| \\
&\leq \left|\epsilon\left(\frac{4\dot{Y}_-^2}{1-Y_-^2} + \frac{4\dot{\theta}_-\dot{Y}_-}{\sqrt{1-Y_-^2}} + 2\dot{\theta}_- O\left(\frac{\epsilon}{\delta}\right) + \frac{4\dot{Y}_-}{\sqrt{1-Y_-^2}} O\left(\frac{\epsilon}{\delta}\right) + O\left(\frac{\epsilon}{\delta}\right)^2\right)\right|.
\end{aligned}$$

By our assumptions, $1 - Y_- \gtrsim \delta$ so it follows that

$$|\dot{Y}_+^2 - \dot{Y}_-^2| \lesssim \delta^2\left(\frac{\delta^2}{\delta} + \frac{\delta}{\sqrt{\delta}} + \delta + \frac{\delta^2}{\sqrt{\delta}} + \delta^2\right) \lesssim \delta^{5/2},$$

which implies $|\dot{Y}_+ - \dot{Y}_-| \lesssim \delta^{3/2}$.

After $n = \lfloor \frac{\mu}{\sqrt{\delta}} \rfloor$ bounces, $|\dot{Y}_n - \dot{Y}_0| \leq \frac{\delta}{2}$ if μ is sufficiently small and we still have the vertical velocity of same order, *i.e.* $\dot{Y}_n \sim \dot{Y}_0 \sim \delta$. Then, at the n^{th} collision, the center of

mass will be located at $y_n \lesssim 1 - \sqrt{\delta}$, which will imply $|\theta_n - \frac{3\pi}{2}| \gtrsim \sqrt[4]{\delta}$. Now using Lemma 5.1, Corollary 5.2, and Theorem 5.3, we compute the error term of the adiabatic invariant under the assumption that the total number of collisions is bounded by $N \lesssim \delta^{-1}$ and the heavy mass does not hit the floor.

$$
\begin{aligned}
&|\dot{\theta}_N| f(Y_N) - |\dot{\theta}_0| f(Y_0) \\
&\qquad = \sum_{k=1}^{n} \left(|\dot{\theta}_k| f(Y_k) - |\dot{\theta}_{k-1}| f(Y_{k-1}) \right) + \sum_{n}^{N} \left(|\dot{\theta}_k| f(Y_k) - |\dot{\theta}_{k-1}| f(Y_{k-1}) \right) \\
&\qquad = \frac{\mu}{\sqrt{\delta}} O\left(\delta\right)) + \left(\frac{C}{\delta} \right) O\left(\delta^{3/2}\right) = O(\sqrt{\delta}).
\end{aligned}
$$

By Theorem 4.4 proved in the previous chapter, there is indeed a uniform bound on the number of bounces.

If the heavy mass does hit the floor it can do so only once as shown in the next section. We claim that the corresponding change in the adiabatic invariant will be only of order δ. Using (3.4) and Remark 3.1 from Chapter 3, we obtain

$$
\begin{aligned}
\dot{Y}_+ &= -\dot{Y}_- + O(\epsilon) \\
\dot{\theta}_+ &= \dot{\theta}_- + O(\epsilon) + O(\delta).
\end{aligned}
$$

Let the pairs $(\dot{\theta}_m, Y_m)$ $(\dot{\theta}_{m+1}, Y_{m+1})$ denote the corresponding values of $(\dot{\theta}, Y)$ when the light mass hits the floor right before and after the large mass hits the floor. Then, since $\dot{Y} = O(\delta)$, we find that $Y_{m+1} - Y_m = O(\delta)$ and $\dot{\theta}_{m+1} - \dot{\theta}_m = O(\delta)$. As a consequence,

$$
|\dot{\theta}_{m+1}| f(Y_{m+1}) - |\dot{\theta}_m| f(Y_m) = O(\delta)
$$

and the change in adiabatic invariant due to large mass hitting the floor is sufficiently small $\Delta I = O(\delta)$. $\square$

CHAPTER 6

Chaotic dynamics of the two-dimensional coin

Using the equal-mass dumbbell model, we study the dynamics of a bouncing coin whose motion is restricted to the two-dimensional plane. As we mentioned in the introduction, the choice of the two-dimensional coin models does not affect our results. To simplify the terminologies, we will refer to the equal-mass dumbbell as a coin.

We first describe the coin system as a point billiard in a transformed domain. Although the majority of the steps overlaps with our discussion from Chapter 3, there are some additional details we need to note.

6.1 The coin system as a gravitational billiard

Consider the coin interacting with the flat boundary under the influence of gravity. Let all other variables be the same as in Chapter 3, except the θ variable. We still define θ to be the angular position of the vector $\overrightarrow{X_{\mathrm{L}}X_{\mathrm{R}}}$ measured counterclockwise from the positive X-direction. However, we measure θ in $\mathbb{R}$ keeping track of the rotation number. We will identify θ and $\theta + 2\pi\mathbb{Z}$ as the same later.

The moment of inertia of the coin is given by $I = m_{\mathrm{L}}(\frac{l}{2})^2 + m_{\mathrm{R}}(\frac{l}{2})^2 = \frac{1}{4}ml^2$. After rescaling, the energy of the system is given by

$$E = \frac{I\dot{\theta}^2}{2} + \frac{m\dot{Y}^2}{2} + mgY = \frac{I\dot{\theta}^2}{2} + \frac{I\dot{y}^2}{2} + g\sqrt{mI}y. \tag{6.1}$$

By considering the corresponding Lagrangian of the system and applying the Hamilton's principle of least action, we know that the motion of the coin between two collisions on

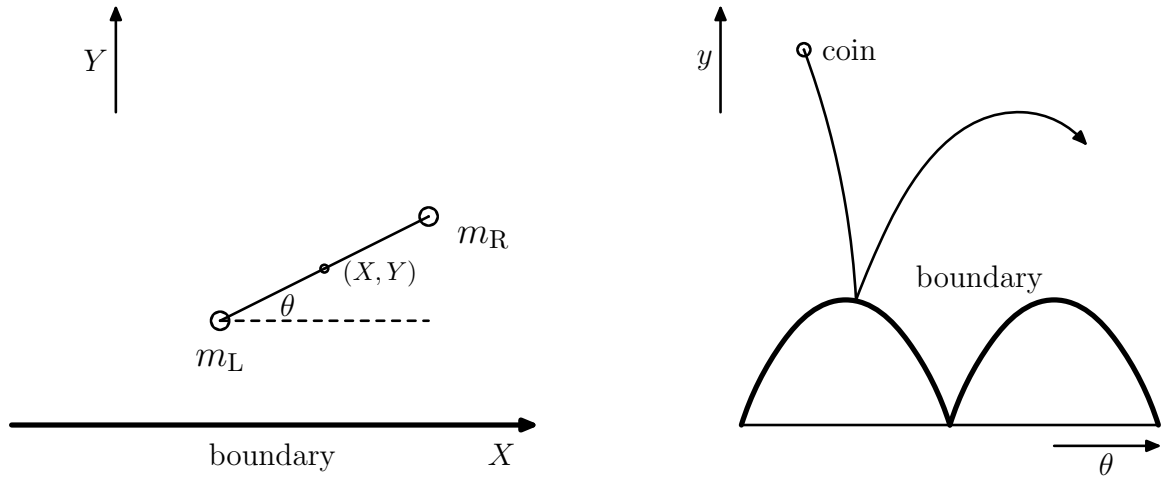


Figure 6.1: (a) The equal-mass dumbbell as a model to the two-dimensional coin. (b) The gravitational billiard for the coin.

the θy plane is governed by

$$\ddot{\theta} = 0 \qquad\qquad \ddot{y} = -g\sqrt{\frac{m}{I}}. \tag{6.2}$$

We choose $m = 1, l = 2$ thus $I = 1$, to simplify calculations.

With the given constants, we find our billiard domain using (3.2),

$$\mathcal{Q} = \{(\theta, y) : y \geq |\sin\theta| \text{ for } \theta \in \mathbb{R}\}. \tag{6.3}$$

At the boundary $\partial\mathcal{Q}$, there will be mirror-like reflections. Applying (3.3) to $\mathbf{v}_- = [\dot{\theta}_-, \dot{y}_-]$ and $\mathbf{n} = \left[1, \cos\theta\frac{\sin\theta}{|\sin\theta|}\right]$, we find the reflection law

$$\begin{pmatrix} \dot{\theta}_+ \\ \dot{y}_+ \end{pmatrix} = \frac{1}{1+\cos^2\theta} \begin{pmatrix} \sin^2\theta & 2\cos\theta\frac{\sin\theta}{|\sin\theta|} \\ 2\cos\theta\frac{\sin\theta}{|\sin\theta|} & -\sin^2\theta \end{pmatrix} \begin{pmatrix} \dot{\theta}_- \\ \dot{y}_- \end{pmatrix}. \tag{6.4}$$

Using (6.2), (6.3), and (6.4), we build a gravitational billiard for the coin system on the θy plane. As in classical billiards, it is sufficient to examine the flow only at the moment of collision. We choose to observe the flow right before a collision. Thus, when there is no ambiguity, we drop the subscript $_-$.

On the natural Poincaré section $\mathcal{P}$

$$\mathcal{P} = \left\{(\theta, y, \dot{\theta}, \dot{y}) : (\theta, y) \in \partial\mathcal{Q},\ \frac{\dot{\theta}^2}{2} + \frac{\dot{y}^2}{2} + gy = E\right\},$$

we define the return map $f : \mathcal{P} \to \mathcal{P}$, which takes the data of a collision to the data of the next collision. A phase point on $\mathcal{P}$ can be identified with (q, v) where q is the footpoint on the boundary $\partial\mathcal{Q}$ and v is the inward velocity vector v at q. If we restrict our attention to the flow pointing "downwards" at the moment of collision,

$$\mathcal{P}_d = \left\{(q, v) : (q, v) \in \mathcal{P}, v \cdot \hat{y} < 0\right\},$$

then we can use $(\theta, \dot{\theta})$ as the coordinates on $\mathcal{P}_d$. Combining (6.1) and (6.3), we check that (q, v) on $\mathcal{P}_d$ can be represented in terms of $(\theta, \dot{\theta})$,

$$q = (\theta, |\sin\theta|) \qquad v = \left(\dot{\theta}, -\sqrt{2E - 2g|\sin\theta| - \dot{\theta}^2}\right). \tag{6.5}$$

When a point and its image under f are both on $\mathcal{P}_d$, we write $f(\theta, \dot{\theta}) = (\theta_1, \dot{\theta}_1)$. In general, given $(\theta, \dot{\theta})$, we cannot solve for $(\theta_1, \dot{\theta}_1)$. However, $(\theta_1, \dot{\theta}_1)$ can be implicitly found in the following way. First, we use (6.4) to write $(\theta_+, \dot{\theta}_+)$ in terms of $(\theta, \dot{\theta})$,

$$\theta_+ = \theta \qquad \dot{\theta}_+ = \frac{1}{1 + \cos^2\theta}\left(\dot{\theta}\sin^2\theta - 2\sqrt{2E - 2g|\sin\theta| - \dot{\theta}^2}\cos\theta\frac{\sin\theta}{|\sin\theta|}\right). \tag{6.6}$$

Given $(\theta_+, \dot{\theta}_+)$, we may use the basic laws of physics to calculate the formula for the parabolic trajectory defined by $(\theta_+, \dot{\theta}_+)$. We may do the same with $(\theta_1, \dot{\theta}_1)$. Since $(\theta_+, \dot{\theta}_+)$ and $(\theta_1, \dot{\theta}_1)$ should define the same parabolic trajectory, we obtain

$$\begin{aligned}
\theta_+ + \frac{\dot{\theta}_+\sqrt{2E - 2g|\sin\theta_+| - \dot{\theta}_+^2}}{g} &= \theta_1 - \frac{\dot{\theta}_-\sqrt{2E - 2g|\sin\theta_1| - \dot{\theta}_1^2}}{g} \\
|\sin\theta_+| + \frac{2E - 2g|\sin\theta_+| - \dot{\theta}_+^2}{2g} &= |\sin\theta_1| + \frac{2E - 2g|\sin\theta_1| - \dot{\theta}_1^2}{2g}.
\end{aligned} \tag{6.7}$$

Combining (6.6) and (6.7) lets us implicitly define $(\theta_1, \dot{\theta}_1)$.

6.2 The topological picture of the billiard map

6.2.1 Construction of the domain D

We will study the topological picture of the return map f which will be fundamental to constructing the horizontal and vertical strips satisfying the Conley-Moser conditions. We start by choosing an appropriate domain for f.

We naturally want to pick a domain which contains the phase points corresponding to collisions of m_{L} and also the phase points corresponding to collisions of m_{R} to study how the two parts interact. From the construction of $\mathcal{P}$, we see that the phase points $(\theta, \dot{\theta})$ for m_{L} satisfy $2n\pi < \theta < (2n+1)\pi$, and the phase points for m_{R} satisfy $(2n-1)\pi < \theta < 2n\pi$. Also, note that $(\theta, \dot{\theta}) \in \mathcal{P}_d$ is not defined when $\theta = n\pi$, since such points are associated to the billiard trajectories hitting the corners of $\partial\mathcal{Q} = |\sin\theta|$. Thus, our domain should not include any points with $\theta = n\pi$. To embrace both conditions, it is apparent we must choose two disjoint subsets as a domain.

Suppose we choose two rectangles of width $2\theta^* < \pi$ and height $2\dot{\theta}^*$ centered at $\left(\frac{\pi}{2}, 0\right)$ and $\left(\frac{3\pi}{2}, 0\right)$ as the domain. As a first step toward understanding the image of the rectangles, let us attempt to calculate the image of one of the corner points, $\left(\frac{\pi}{2} - \theta^*, \dot{\theta}^*\right)$. To find its image, we first need to know how $\dot{\theta}^*$ changes right after the reflection. By (6.6),

$$\dot{\theta}^*_+ = \frac{1}{1 + \sin^2\theta^*}\left((1 - \sin^2\theta^*)\dot{\theta}^* - 2\sin\theta^*\sqrt{2E - 2g\cos\theta^* - (\dot{\theta}^*)^2}\right).$$

We can imagine that using the exact form of this equation in the next steps will cause complicated calculations. However, if θ^* is small and $(\dot{\theta}^*)^2$ is small compared to $2E$, then we can easily estimate the θ^* and $\dot{\theta}^*$, thus the equation as well, using asymptotics.

We assume the energy of the system E is large and set $\theta^* = O\left(\frac{1}{E}\right)$ and $\dot{\theta}^* = O\left(\frac{1}{\sqrt{E}}\right)$. We note that, with the condition $\theta^* = O\left(\frac{1}{E}\right)$, we only consider the pieces of the boundary

$\partial\mathcal{Q}$ "near" the sine peaks. The restriction in $\dot{\theta}^*$ implies that the angle between the incoming billiard trajectory and $-\hat{y}$ at the moment of a collision is $O\left(\frac{1}{E}\right)$, i.e. the trajectory is "close to vertical". Therefore, this setting guarantees that we will always have the transversal Poincaré section.

Let $D = D(E, k(E))$ be the union of

$$D_{\mathrm{L}} = \left\{(\theta, \dot{\theta}) : (\theta, \dot{\theta}) \in \mathcal{P}_d, \left|\theta - \frac{1}{2}\pi\right| \leq \frac{k}{E}, |\dot{\theta}| \leq \frac{\sqrt{2}k}{\sqrt{E}}\right\}$$
$$D_{\mathrm{R}} = \left\{(\theta, \dot{\theta}) : (\theta, \dot{\theta}) \in \mathcal{P}_d, \left|\theta - \frac{3}{2}\pi\right| \leq \frac{k}{E}, |\dot{\theta}| \leq \frac{\sqrt{2}k}{\sqrt{E}}\right\},$$

where k is a function of E satisfying the equation (6.8) in Lemma 6.2 below, and k is bounded by two independent constants $0 < k_1 < k(E) < k_2$.

Remark 6.1. We explain why we choose such specific $k(E)$. As k changes, the topological picture of f changes in the following way.

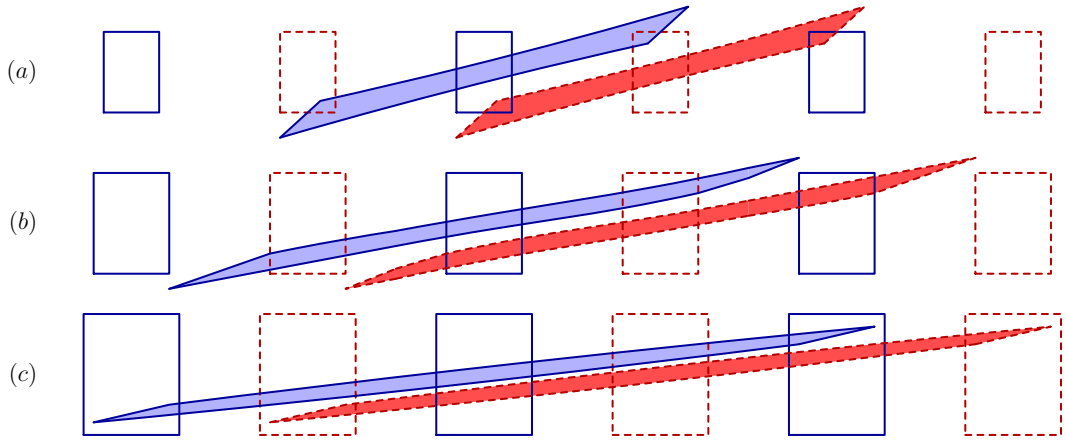


Figure 6.2: (a) When $K<k$. (b) When using k defined in Lemma 6.2. (c) When $K>k$.

1. If we instead use a constant K slightly smaller than k when defining the domain $D = D(E, K)$, then $f(D(E, K))$ intersects $D(E, K)\pm(2\pi\mathbb{Z}, 0)$ in two horizontal strips and four corners (Figure 6.2a).

2. As K approaches k, $f(D(E, K))$ becomes wider and flatter. When we use the exact k, then $f(D(E, k)) \cap (D(E, k)\pm(2\pi\mathbb{Z}, 0))$ results in six horizontal strips as in Figure

6.2b. These six horizontal strips are the "minimal" number of strips such that, later when we identify the angle of the coin $\theta = \theta + 2\pi\mathbb{Z}$, the image of each rectangle intersects itself "and" the other rectangle in full horizontal strips (Figure 6.5a).

3. With K slightly bigger than k, $f(D(E,K))$ still intersects in $(D(E,K)\pm(2\pi\mathbb{Z},0))$ in six horizontal strips as in the case of k, but it is more tedious to compute the coordinates of $f(D(E,K))$.

4. As K gets bigger, the topological picture of $f(D(E,K))$ changes from Figure 6.2b to Figure 6.2c.

Lemma 6.2. *Suppose we have a large enough energy of the system E. Then there exists $k = k(E)$ such that given a point $a = (\theta^a, \dot\theta^a) = \left(\frac{\pi}{2} - \frac{k}{E}, \frac{\sqrt{2}k}{\sqrt{E}}\right) \in \mathcal{P}_d$, we have*

$$A(E,k) = \frac{2}{g}\left(\dot\theta^a_+ \sqrt{2E - 2g|\sin\theta^a| - \left(\dot\theta^a_+\right)^2}\right) + \pi = 0. \tag{6.8}$$

Remark 6.3. The equation $A(E,k) = 0$ implies the following physical meaning: the billiard trajectory on $\mathcal{Q}$ defined by the initial condition $a = (\theta^a, \dot\theta^a)$ hits the boundary again when $\theta = \theta^a - \pi$ (Figure 6.4). In the phase space, this guarantees that a corner point $a \in D_{\mathrm{L}}$ gets mapped to the left side of $D_{\mathrm{R}} - (2\pi, 0)$ (Figure 6.3).

Proof. Recall from (6.5) that when $\dot\theta^a = \frac{\sqrt{2}k}{\sqrt{E}}$, we have $\dot y^a = -\sqrt{2E - 2g|\sin\theta^a| - (\dot\theta^a)^2}$. By (6.6),

$$\begin{aligned}
\dot\theta^a_+ &= \frac{1}{1+\sin^2\frac{k}{E}}\left(\left(1-\sin^2\frac{k}{E}\right)\frac{\sqrt{2}k}{\sqrt{E}} - 2\sin\frac{k}{E}\sqrt{2E - 2g\cos\frac{k}{E} - \left(\sqrt{2E}\frac{k}{E}\right)^2}\right) \\
&= \frac{1}{1+\sin^2\frac{k}{E}}\left(\left(1 - O\left(\frac{k^2}{E^2}\right)\right)\frac{\sqrt{2}k}{\sqrt{E}} - \left(\frac{2k}{E} - O\left(\frac{k^3}{E^3}\right)\right)\left(\sqrt{2E} - O\left(\frac{1}{E^{0.5}}\right)\right)\right) \\
&= \frac{1}{1+\sin^2\frac{k}{E}}\left(-\frac{\sqrt{2}k}{\sqrt{E}} + O\left(\frac{k}{E^{1.5}}\right)\right) = -\frac{\sqrt{2}k}{\sqrt{E}} + O\left(\frac{k}{E^{1.5}}\right).
\end{aligned} \tag{6.9}$$

Then we substitute (6.9) to (6.8)

$$\begin{aligned}
A(E,k) &= \frac{2}{g}\left(\left(-\frac{\sqrt{2}k}{\sqrt{E}}+O\left(\frac{k}{E^{1.5}}\right)\right)\sqrt{2E-2g\cos\frac{k}{E}-\left(-\frac{\sqrt{2}k}{\sqrt{E}}+O\left(\frac{k}{E^{1.5}}\right)\right)^2}\right)+\pi \\
&= \frac{2}{g}\left(\left(-\frac{\sqrt{2}k}{\sqrt{E}}+O\left(\frac{k}{E^{1.5}}\right)\right)\left(\sqrt{2E}-O\left(\frac{1}{E^{1.5}}\right)\right)\right)+\pi \\
&= \frac{2}{g}\left(-2k+O\left(\frac{k}{E}\right)\right)+\pi.
\end{aligned}$$

Since the term $O\left(\frac{k}{E}\right)$ is small, we may assume $A\left(E,\frac{g\pi}{2}\right)<0$ and $A\left(E,\frac{g\pi}{8}\right)>0$. By the intermediate value theorem, we can find $k\in\left(\frac{g\pi}{8},\frac{g\pi}{2}\right)$ such that $A(E,k)=0$. Since the leading term of $\frac{\partial A(E,k)}{\partial k}=-\frac{2}{g}$, it follows that $\frac{\partial A(E,k)}{\partial k}\neq 0$. From the implicit function theorem, we know that we can write $k=k(E)$ in terms of E. □

6.2.2 The image of D under the billiard map

We will use a few reference points in D and the Jacobian approximation of f to show that $f(D)$ are long and thin strips overlapping D and $D\pm(2\pi,0)$ as in Figure 6.3. Due to the symmetry of the coin, $f(D_{\mathrm{R}})$ will have the same topological picture as $f(D_{\mathrm{L}})$, shifted by π. Therefore, we may focus on the topological behavior of $f(D_{\mathrm{L}})$. We start by analyzing the image of the top edge of D_{L} under f.

Lemma 6.4. *The image of the top edge of D_{L} under f is a μ_h-horizontal curve with $\mu_h=O\left(\frac{1}{\sqrt{E}}\right)$.*

Proof. We first compute the Jacobian $\mathbf{J}_f$ for general $(\theta,\dot\theta)$, then estimate the transformation of a vector $(1,0)^T$ in D_{L} by $\mathbf{J}_f$. From (6.6), we get the derivatives

$$\begin{aligned}
\frac{\partial\theta_+}{\partial\theta}&=1 \qquad & \frac{\partial\dot\theta_+}{\partial\theta}&=\frac{-g\cos^2\theta(3+\cos 2\theta)+2\dot\theta\dot y\sin 2\theta-2\dot y^2\sin^3\theta}{\dot y(1+\cos^2\theta)^2}\\
\frac{\partial\theta_+}{\partial\dot\theta}&=0 \qquad & \frac{\partial\dot\theta_+}{\partial\dot\theta}&=\frac{1-\cos^2\theta-(2\dot\theta\cos\theta)/\dot y}{1+\cos^2\theta}.
\end{aligned}$$

By implicitly differentiating (6.7) and simplifying the results with

$$\dot{y} = -\sqrt{2E - 2g\sin\theta - \dot{\theta}^2} \qquad \dot{y}_+ = \sqrt{2E - 2g\sin\theta - \dot{\theta}_+^2} \qquad \dot{y}_1 = -\sqrt{2E - 2g|\sin\theta_1| - (\dot{\theta}_1)^2},$$

we obtain

$$\begin{aligned}
\left(-1 \pm \frac{\dot{\theta}_1\cos\theta_1}{\dot{y}_1}\right)\frac{\partial\theta_1}{\partial\theta_+} - \left(\frac{(\dot{y}_1)^2 - (\dot{\theta}_1)^2}{g\dot{y}_1}\right)\frac{\partial\dot{\theta}_1}{\partial\theta_+} &= \frac{\dot{\theta}_+\cos\theta_+}{\dot{y}_+} - 1 \\
\frac{\theta_1}{g}\ \frac{\partial\dot{\theta}_1}{\partial\theta_+} &= 0 \\
\left(-1 \pm \frac{\dot{\theta}_1\cos\theta_1}{\dot{y}_1}\right)\frac{\partial\theta_1}{\partial\dot{\theta}_+} - \left(\frac{(\dot{y}_1)^2 - (\dot{\theta}_1)^2}{g\dot{y}_1}\right)\frac{\partial\dot{\theta}_1}{\partial\dot{\theta}_+} &= -\frac{(\dot{y}_+)^2 - (\dot{\theta}_+)^2}{g\dot{y}_+} \\
-\frac{\dot{\theta}_1}{g}\ \frac{\partial\dot{\theta}_1}{\partial\dot{\theta}_+} &= -\frac{\dot{\theta}_+}{g},
\end{aligned}$$

where the $\pm$ signs depend on the sign of $\frac{\sin\theta_1}{|\sin\theta_1|}$. Keeping in mind that $\dot{\theta}_+ = \dot{\theta}_1$, we solve for

$$\frac{\partial\theta_1}{\partial\theta_+} = \frac{\dot{y}_1(\dot{y}_+ - \dot{\theta}_+\cos\theta_+)}{\dot{y}_+(\dot{y}_1 \mp \dot{\theta}_1\cos\theta_1)} \qquad \frac{\partial\dot{\theta}_1}{\partial\theta_+} = 0 \qquad \frac{\partial\theta_1}{\partial\dot{\theta}_+} = \frac{(\dot{y}_+ - \dot{y}_1)(\dot{y}_+\dot{y}_1 + (\dot{\theta}_1)^2)}{g\dot{y}_+(\dot{y}_1 \mp \dot{\theta}_1\cos\theta_1)} \qquad \frac{\partial\dot{\theta}_1}{\partial\dot{\theta}_+} = 1.$$

We use the chain rule to compute

$$\mathbf{J}_f = \begin{bmatrix} \dfrac{\partial\theta_1}{\partial\theta} & \dfrac{\partial\theta_1}{\partial\dot{\theta}} \\ \dfrac{\partial\dot{\theta}_1}{\partial\theta} & \dfrac{\partial\dot{\theta}_1}{\partial\dot{\theta}} \end{bmatrix} = \begin{bmatrix} \dfrac{\partial\theta_1}{\partial\theta_+}\dfrac{\partial\theta_+}{\partial\theta} + \dfrac{\partial\theta_1}{\partial\dot{\theta}_+}\dfrac{\partial\dot{\theta}_+}{\partial\theta} & \dfrac{\partial\theta_1}{\partial\theta_+}\dfrac{\partial\theta_+}{\partial\dot{\theta}} + \dfrac{\partial\theta_1}{\partial\dot{\theta}_+}\dfrac{\partial\dot{\theta}_+}{\partial\dot{\theta}} \\ \dfrac{\partial\dot{\theta}_1}{\partial\theta_+}\dfrac{\partial\theta_+}{\partial\theta} + \dfrac{\partial\dot{\theta}_1}{\partial\dot{\theta}_+}\dfrac{\partial\dot{\theta}_+}{\partial\theta} & \dfrac{\partial\dot{\theta}_1}{\partial\theta_+}\dfrac{\partial\theta_+}{\partial\dot{\theta}} + \dfrac{\partial\dot{\theta}_1}{\partial\dot{\theta}_+}\dfrac{\partial\dot{\theta}_+}{\partial\dot{\theta}} \end{bmatrix}.$$

For $(\theta, \dot{\theta})$ in D_{L},

$$\begin{array}{rrrr}
\dot{\theta} = \pm O\left(\frac{1}{\sqrt{E}}\right) & \dot{y}_+ = & O(\sqrt{E}) & \theta = \frac{\pi}{2} \pm O\left(\frac{1}{E}\right) \\
\dot{\theta}_+, \dot{\theta}_1 = \pm O\left(\frac{1}{\sqrt{E}}\right) & \dot{y}, \dot{y}_1 = & -O(\sqrt{E}) & \theta_1 = O(1).
\end{array} \tag{6.10}$$

Using (6.10), we estimate the Jacobian $\mathbf{J}_f$ for the points in D_{L}

$$(\mathbf{J}_f)|_{D_{\mathrm{L}}} = \begin{bmatrix} O(1)\cdot 1 + O(\sqrt{E})O(\sqrt{E}) & O(1)\cdot 0 + O(\sqrt{E})\cdot 1 \\ 0\cdot 1 + 1\cdot O(\sqrt{E}) & 0\cdot 0 + 1\cdot O(1) \end{bmatrix} = \begin{bmatrix} O(E) & O(\sqrt{E}) \\ O(\sqrt{E}) & O(1) \end{bmatrix}. \tag{6.11}$$

Any vector $(1,0)^T$ in D_{L} is mapped to $(\mathbf{J}_f)|_{D_{\mathrm{L}}}(1,0)^T = (O(E), O(\sqrt{E}))^T$. It is clear that the image of the top edge of D_{L} is a μ_h-horizontal curve with $\mu_h = O\big(\frac{1}{\sqrt{E}}\big)$. Moreover, since we were careful with the signs of the asymptotic terms, it is a monotonically increasing μ_h-horizontal curve with $\mu_h = O\big(\frac{1}{\sqrt{E}}\big)$. □

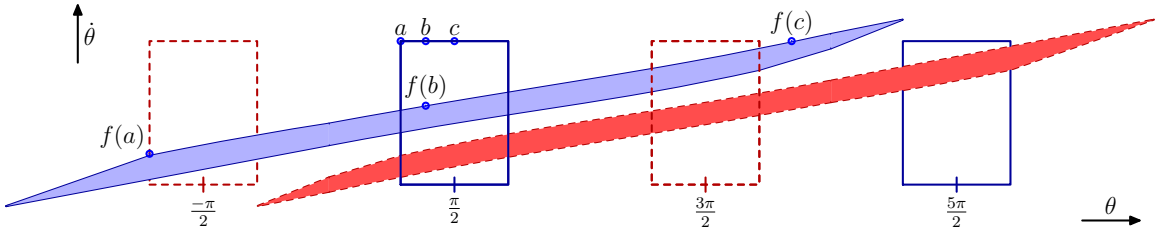


Figure 6.3: A topological picture of $f(D_{\mathrm{L}} \cup D_{\mathrm{R}})$.

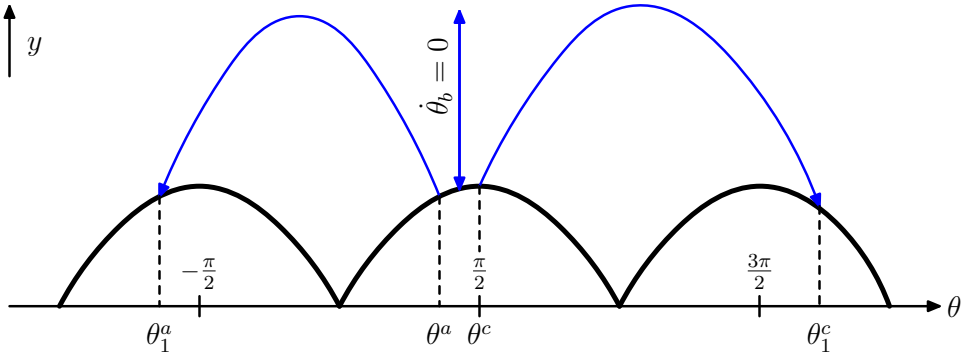


Figure 6.4: The trajectories defined by the initial points a, b, and c on $\mathcal{Q}$.

Now we consider the three reference points a, b, and c lying on the top edge of D_{L} (see Figure 6.3),

$$a = (\theta^a, \dot\theta^a) = \Big(\frac{\pi}{2} - \frac{k}{E}, \frac{\sqrt{2}k}{\sqrt{E}}\Big) \qquad b = (\theta^b, \dot\theta^b) = \Big(\theta^b, \frac{\sqrt{2}k}{\sqrt{E}}\Big) \qquad c = (\theta^c, \dot\theta^c) = \Big(\frac{\pi}{2}, \frac{\sqrt{2}k}{\sqrt{E}}\Big),$$

where θ^b is chosen such that $\dot{\theta}^b_+ = \dot{\theta}^b_1 = 0$. The existence of such θ^b can be shown by applying the intermediate value theorem to the function $\dot{\theta}_1\left(\theta, \sqrt{2E}\frac{k}{E}\right)$. See also Figure 6.4 for a geometric reason.

Proposition 6.5. *The image of D overlaps $D \pm (2\pi\mathbb{Z}, 0)$ in six disjoint μ_h-horizontal strips with $\mu_h = O\left(\frac{1}{\sqrt{E}}\right)$ (Figure 6.3).*

Proof. We estimate the images of the three reference points a, b, and c in D_{L}

$$f(a) = (\theta^a_1, \dot{\theta}^a_1) \qquad f(b) = (\theta^b_1, \dot{\theta}^b_1) \qquad f(c) = (\theta^c_1, \dot{\theta}^c_1).$$

From (6.9), we know $\dot{\theta}^a_1 = \dot{\theta}^a_+ = -\frac{\sqrt{2}k}{\sqrt{E}} + O\left(\frac{k}{E^{1.5}}\right)$. To find θ^1_a, we use the first term of $A(E, k)$ from Lemma 6.2. The term represents twice the θ-distance from θ_a to the vertex of the parabolic billiard trajectory determined by $(\theta^a, \dot{\theta}^a)$, and it is set to $-\pi$. In other words, the next collision occurs when

$$\theta^1_a = \theta_a - \pi = -\frac{\pi}{2} - \frac{k}{E}.$$

By construction,

$$f(b) = (\theta^b_1, \dot{\theta}^b_1) = (\theta^b, 0).$$

Since the point c is associated to the trajectory hitting the boundary at the peak (Figure 6), the $\dot{\theta}$-component does not change after the collision, and $\dot{\theta}^c_1 = \sqrt{2E}\frac{k}{E}$. To estimate θ^c_1, we observe that the parabolic billiard trajectory determined by $(\theta^c, \dot{\theta}^c)$ has the footpoint $q^c = (\theta^c, y^c) = (\frac{\pi}{2}, 1)$ on $\partial\mathcal{Q}$. Any footpoint on $\partial\mathcal{Q}$, including the footpoint for the next collision $q^c_1 = (\theta^c_1, y^c_1)$, has the y-coordinate less than or equal to 1. This implies that y- and θ-distances from q^c to the vertex of the parabola is smaller than the y- and θ-distances from q^c_1 to the vertex of the parabola. In other words,

$$\theta^c_1 \geq \theta^c + \frac{2}{g}\left(\dot{\theta}^c_+\sqrt{2E - 2g|\sin\theta^c| - (\dot{\theta}^c_+)^2}\right). \tag{6.12}$$

We use (6.9) to get the relation $\dot{\theta}^c = \frac{\sqrt{2}k}{\sqrt{E}} = |\dot{\theta}^a| + O\big(\frac{k}{E^{1.5}}\big)$. Then we rewrite (6.12) as

$$\begin{aligned}
\theta_1^c &\geq \frac{\pi}{2} + \frac{2}{g}\left(\Big(|\dot{\theta}_+^a| + O\Big(\frac{k}{E^{1.5}}\Big)\Big)\sqrt{2E - 2g\cos\frac{k}{E} - 2g\Big(1-\cos\frac{k}{E}\Big) - \Big(|\dot{\theta}_+^a| + O\Big(\frac{k}{E^{1.5}}\Big)\Big)^2}\right) \\
&= \frac{\pi}{2} + \frac{2}{g}\left(\Big(|\dot{\theta}_+^a| + O\Big(\frac{k}{E^{1.5}}\Big)\Big)\sqrt{2E - 2g\sin\theta^a - (\dot{\theta}_+^a)^2 - 2g\Big(1-\cos\frac{k}{E}\Big) - O\Big(\frac{k^2}{E^2}\Big)}\right) \\
&= \frac{\pi}{2} + \frac{2}{g}\left(\Big(|\dot{\theta}^a| + O\Big(\frac{k}{E^{1.5}}\Big)\Big)\left(\sqrt{2E - 2g\sin\theta^a - (\dot{\theta}_+^a)^2)} - O\Big(\frac{1}{E^{0.5}}\Big)\right)\right) \\
&= \frac{\pi}{2} + \frac{2}{g}\left(|\dot{\theta}^a|\sqrt{2E - 2g\sin\theta^a - (\dot{\theta}_+^a)^2)}\right) + \text{smaller terms},
\end{aligned}$$

where

$$\text{smaller terms} = \frac{2}{g}\left(-|\dot{\theta}^a| O\Big(\frac{1}{E^{1.5}}\Big) + O\Big(\frac{k}{E^{1.5}}\Big)\sqrt{2E - 2g\sin\theta^a - (\dot{\theta}_+^a)^2} - O\Big(\frac{k}{E^2}\Big)\right). \tag{6.13}$$

It can be shown that (6.13) is $O\big(\frac{k}{E}\big)$ and is greater than or equal to $\frac{k}{E}$. Since we know from (6.8) that $\frac{2}{g}(|\dot{\theta}^a|\sqrt{2E - 2g\sin\theta^a - (\dot{\theta}_+^a)^2)}) = A(E,k) = \pi$, we conclude that $\theta_c = \frac{3\pi}{2} + \frac{k}{E} + O\big(\frac{k}{E}\big)$.

We plot the images of the three reference points $f(a), f(b)$, and $f(c)$ (Figure 6.3). Then by Lemma 6.4, we know that the image of the top side of D_{L} horizontally crosses $D_{\mathrm{L}}, D_{\mathrm{R}}$ and $D_{\mathrm{R}} - (2\pi, 0)$. Similarly, using symmetries, we infer that the image of the bottom edge of D_{L} also crosses $D_{\mathrm{L}}, D_{\mathrm{R}}$ and $D_{\mathrm{R}} - (2\pi, 0)$ horizontally. Since f is continuous and the proof of Lemma 6.4 can be applied to any vector $(1,0)^T$ lying in D_{L}, we deduce that the image of D_{L} is μ_h-horizontal strip crossing $D_{\mathrm{L}}, D_{\mathrm{R}}$ and $D_{\mathrm{R}} - (2\pi, 0)$ horizontally. The same argument works for D_{R}, and we conclude that $f(D)$ on $D \pm (2\pi\mathbb{Z}, 0)$ are six horizontal strips. $\square$

6.3 Construction of the strips satisfying the Conley-Moser conditions

In this section, we finally construct the horizontal and vertical strips satisfying the Conley-Moser conditions using the coin billiard map. Recall that we measured the angular position of the coin $\theta \in \mathbb{R}$ distinguishing θ from $\theta + 2\pi\mathbb{Z}$. We now impose the equivalence relation $\theta \sim \theta + 2\pi\mathbb{Z}$ and treat $\theta = \theta + 2\pi\mathbb{Z}$ as the same. We denote the resultant objects with $\widetilde{\ }$, i.e. $\widetilde{\mathcal{Q}}, \widetilde{\mathcal{P}_d}, \widetilde{D}, \tilde{f}$.

Under this setting where $\theta \in [0, 2\pi)$, $\widetilde{D}_{\mathrm{L}}$ and $\widetilde{D}_{\mathrm{R}}$ lie on a cylinder. It directly follows from Proposition 6.5 and Figure 6.3 that

Corollary 6.6. *$\tilde{f}(\widetilde{D}) \cap \widetilde{D}$ are six μ_h-horizontal strips with $\mu_h = O\left(\frac{1}{\sqrt{E}}\right)$ (Figure 6.5a).*

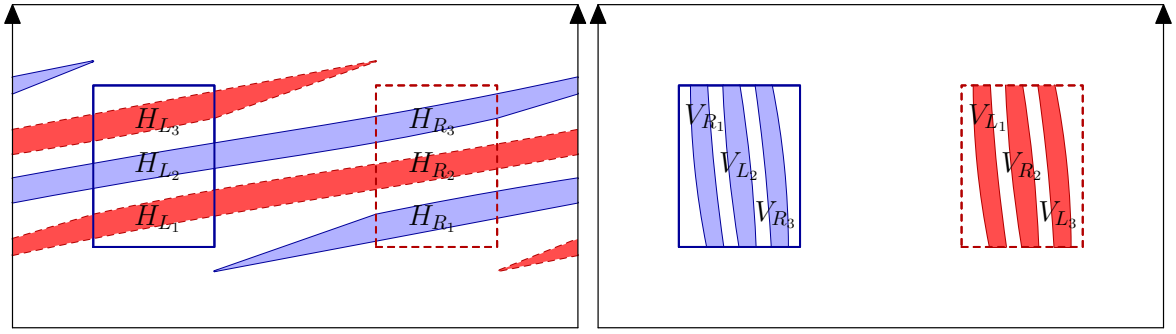


Figure 6.5: (a) The two rectangles $\widetilde{D}_{\mathrm{L}}$ (solid) and $\widetilde{D}_{\mathrm{R}}$ (dashed) and their first images under the billiard map $\tilde{f}$. (b) Vertical strips V_s which are the inverse images of $\widetilde{D}_{\mathrm{L}}$ and $\widetilde{D}_{\mathrm{R}}$.

We denote the six horizontal strips H_s where $s \in S = \{\mathrm{L}_1, \mathrm{L}_2, \mathrm{L}_3, \mathrm{R}_1, \mathrm{R}_2, \mathrm{R}_3\}$ as illustrated in Figure 6.5a.

Proposition 6.7. *The preimages of H_s are six disjoint μ_v-vertical strips $V_s \in \widetilde{D}$ with $\mu_v = O\left(\frac{1}{\sqrt{E}}\right)$.*

Proof. Without loss of generality, we choose H_{L_2} to find its preimage. From the construction of the horizontal strips H_s, the inverse images of the upper and lower boundary curves of H_{L_2} are segments of the top and the bottom edges of $\widetilde{D}_{\mathrm{L}}$. To find the inverse image of the vertical boundaries of H_{L_2}, we compute $\mathbf{J}_{\tilde{f}^{-1}}(0,1)^T$. By the inverse function theorem,

$\mathbf{J}_{\tilde{f}^{-1}} = (\mathbf{J}_{\tilde{f}})^{-1}$. Since f and $\tilde{f}$ are locally the same, we may use (6.11) and the equations in the proof of Lemma 6.4 to compute $\mathbf{J}_{\tilde{f}^{-1}}$. First, by direct computation, we find the determinant of $\mathbf{J}_{\tilde{f}}$

$$|\mathbf{J}_{\tilde{f}}| = \frac{\dot{y}_1(\dot{y}_+ - \dot{\theta}_1 \cos\theta)(-2\dot{\theta}\cos\theta + \dot{y}\sin^2\theta)}{\dot{y}\dot{y}_+(1+\cos^2\theta)(\dot{y}_1 \mp \dot{\theta}_1\cos\theta_1)}.$$

Using (6.10), we find $|\mathbf{J}_{\tilde{f}}| = O(1)$ for the points in $\widetilde{D}_{\mathrm{L}}$. Thus,

$$(\mathbf{J}_{\tilde{f}^{-1}})|_{H_{\mathrm{L}_2}} = \frac{1}{O(1)} \begin{bmatrix} O(1) & -O(\sqrt{E}) \\ -O(\sqrt{E}) & O(E) \end{bmatrix},$$

and $(\mathbf{J}_{\tilde{f}^{-1}})|_{H_{L_2}}(0,1)^T = (-O(\sqrt{E}), O(E))^T$. We then obtain $\mu_v = O\big(\frac{1}{\sqrt{E}}\big)$.

To summarize, the preimages of the boundaries of H_{L_2} consist of two segments of the top and the bottom edges of $\widetilde{D}_{\mathrm{L}}$ and two μ_v-vertical curves in $\widetilde{D}_{\mathrm{L}}$. Since $\tilde{f}^{-1}$ is a diffeomorphism, the boundaries get mapped to the boundaries and the interior gets mapped to the interior. Therefore the inverse image of a horizontal strip H_{L_2} is a vertical strip.

□

Proposition 6.8.

1. *Consider a μ_h-horizontal strip H_s for $s \in \{\mathrm{L}_1, \mathrm{L}_2, \mathrm{L}_3\}$. If $j \in \{\mathrm{R}_1, \mathrm{L}_2, \mathrm{R}_3\}$, then $\tilde{f}(H_s) \cap H_j$ is a μ_h^1-horizontal strip where $\mu_h^1 = O\big(\frac{1}{\sqrt{E}}\big)$ and $d(\tilde{f}(H_s) \cap H_j) < d(H_s)$. If $j \in \{\mathrm{L}_1, \mathrm{R}_2, \mathrm{L}_3\}$, then $\tilde{f}(H_s) \cap H_j = \emptyset$.*

2. *Similary, consider a μ_v-vertical strip V_s for $s \in \{\mathrm{R}_1, \mathrm{L}_2, \mathrm{R}_3\}$. If $j \in \{\mathrm{L}_1, \mathrm{L}_2, \mathrm{L}_3\}$, then $\tilde{f}^{-1}(V_s) \cap V_j$ is a μ_v^1-vertical where $\mu_v^1 = O\big(\frac{1}{\sqrt{E}}\big)$ and $d(\tilde{f}^{-1}(V_s) \cap V_j) < d(V_s)$. If $j \in \{\mathrm{R}_1, \mathrm{R}_2, \mathrm{R}_3\}$, then $\tilde{f}^{-1}(V_s) \cap V_j = \emptyset$.*

3. *The statements 1 and 2 with L and R switched are also true.*

Proof. We only prove the statement 1, since the proofs for the statements 2 and 3 are similar. By observing the topological pictures of $\tilde{f}(\widetilde{D})$ (Figure 6.3, Figure 6.5a), we can easily see that when $s \in \{\mathrm{L}_1, \mathrm{L}_2, \mathrm{L}_3\}$ and $j \in \{\mathrm{L}_1, \mathrm{R}_2, \mathrm{L}_3\}$, we have $\tilde{f}(H_s) \cap H_j = \emptyset$. If $j \in \{\mathrm{L}_1, \mathrm{R}_2, \mathrm{L}_3\}$,

then $\tilde{f}(H_s) \cap H_j$ is nonempty and is contained in H_j. It is clear $d(\tilde{f}(H_s) \cap H_j) \leq d(H_s)$. It remains to estimate the maximal slope of the horizontal boundary curves of $\tilde{f}(H_s) \cap H_j$. From Proposition 6.5, we know that the horizontal boundary curves of H_s has the maximal slope $\mu_h = O\big(\frac{1}{\sqrt{E}}\big)$. Using (6.11), we get

$$(\mathbf{J}_{\tilde{f}})|_D \begin{pmatrix} 1 \\ \mu_h \end{pmatrix} = \begin{pmatrix} O(E) + O(\sqrt{E})O\big(\frac{1}{\sqrt{E}}\big) \\ O(\sqrt{E}) + O(\frac{1}{\sqrt{E}}) \end{pmatrix} = \begin{pmatrix} O(E) \\ O(\sqrt{E}) \end{pmatrix}.$$

Thus, the maximal slope of the horizontal boundary curves of $f(H_s) \cap H_j$ is $\mu_h^1 = \frac{O(\sqrt{E})}{O(E)} = O\big(\frac{1}{\sqrt{E}}\big)$.

□

Corollary 6.6, Proposition 6.7, and Proposition 6.8 together imply that $\tilde{f}$ satisfies the modified Conley-Moser conditions. We state our main theorem, which is merely a version of Theorem 2.4. We first need a few definitions.

Definition. Let $\Lambda_N = \{p \mid \bigcap_{n=-N}^{N} \tilde{f}^n(\widetilde{D})\}$ and $\Lambda = \lim_{N\to\infty} \Lambda_N$.

Definition. Let Σ be the set of bi-infinite sequences of six symbols $\mathrm{L}_1, \mathrm{L}_2, \mathrm{L}_3, \mathrm{R}_1, \mathrm{R}_2, \mathrm{R}_3$ with the following rules:

- $\mathrm{L}_1, \mathrm{L}_2$, or L_3 can precede $\mathrm{R}_1, \mathrm{L}_2$, or R_3.
- $\mathrm{R}_1, \mathrm{R}_2$, or R_3 can precede $\mathrm{L}_1, \mathrm{R}_2$, or L_3.

Theorem 6.9. *There is a homeomorphism $\phi : \Lambda \to \Sigma$ such that if we denote the shift map on Σ as $\sigma : \Sigma \to \Sigma$, then the diagram below commutes.*

$$\begin{array}{ccc} \Lambda & \xrightarrow{\tilde{f}} & \Lambda \\ \phi \downarrow & & \downarrow \phi \\ \Sigma & \xrightarrow[\sigma]{} & \Sigma \end{array}$$

Proof. The proof closely follows the proof of Theorem 2.4. We only point out some key ideas here.

We first describe the invariant set Λ: the set of points which stay in $\widetilde{D}$ after infinitely many forward and backward iterations of $\tilde{f}$. From Corollary 6.6 and Figure 6.5a, we see that after one forward iteration of $\tilde{f}$, the invariant set $\tilde{f}(\widetilde{D}) \cap \widetilde{D}$ consists of the $2 \cdot 3$ horizontal strips H_s where $s \in S$. From Proposition 6.8 and Figure 6.5a, we know that $\tilde{f}$ acting on H_s creates 3 nested μ_h^1-horizontal strips with $\mu_h^1 = O\big(\frac{1}{\sqrt{E}}\big)$ in each H_s. In other words,

$$\tilde{f}^2(\widetilde{D}) \cap \tilde{f}(\widetilde{D}) \cap \widetilde{D} = \tilde{f}(H_s) \cap H_j \text{ where } s, j \in S$$

consists of $2 \cdot 3^2$ thinner μ_h^1-horizontal strips. We define each nested strip

$$\tilde{f}(H_{s_{-2}}) \cap H_{s_{-1}} = H_{s_{-2}s_{-1}} \text{ where } s_i \in S.$$

If we iterate $\tilde{f}^n$, then $2 \cdot 3^n \mu_h^n$-horizontal strips with $\mu_h^n = O\big(\frac{1}{\sqrt{E}}\big)$ remain in $\widetilde{D}$. We inductively define the nested horizontal strips after n iterations,

$$\tilde{f}(H_{s_{-n}\dots s_{-2}}) \cap H_{s_{-1}} = H_{s_{-n}\dots s_{-1}}.$$

Similarly, from Propositions 6.7 and 6.8, we obtain $2 \cdot 3^n$ μ_v^*-vertical strips with $\mu_v^n = O\big(\frac{1}{\sqrt{E}}\big)$ as the invariant set for $\tilde{f}^{-n}$.

We see that Λ_N is the intersection of $2 \cdot 3^N$ μ_h^N-horizontal strips and $2 \cdot 3^N$ μ_v^N-vertical strips. When $n \to \infty$, we see that Λ is a Cantor set which contains all the intersection points of an infinite number of μ_h^*-horizontal curves and μ_v^*-vertical curves. The first Conley-Moser condition $0 < \mu_h \mu_v = O(\frac{1}{\sqrt{E}})O\big(\frac{1}{\sqrt{E}}\big) < 1$ guarantees that there exists a unique intersection point of a μ_h-horizontal curve and a μ_v-vertical curve.

The next step is to assign a sequence $s = (\dots s_{-2}s_{-1}s_0s_1\dots) \in \Sigma$ to each point $p \in \Lambda$ based on to which horizontal curve $H_{\dots s_{-2}s-1}$ and vertical curve $V_{s_0s_1\dots}$ the point p belongs. It is clear from the construction that this assignment $\phi : \Sigma \to \Lambda$ is one-to-one. Also, from Proposition 6.8, we know that only certain choices of s, j for $f(H_s) \cap H_j = H_{sj}$ are valid.

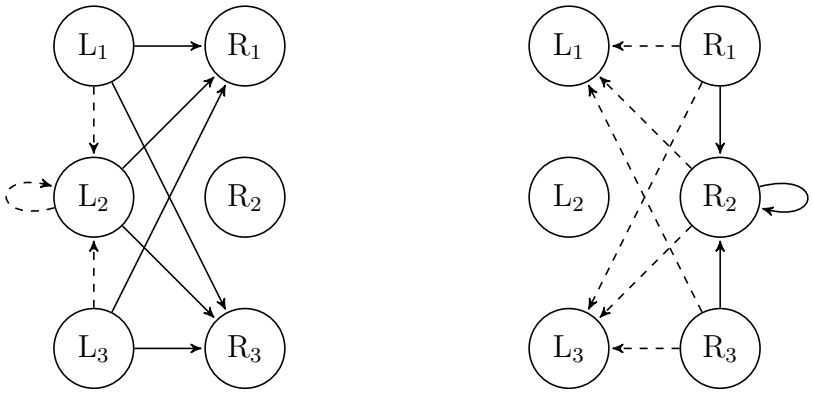


Figure 6.6: The valid letter combinations for the sequences in Σ.

Thus, the sequences $s \in \Sigma$ must combinatorially follow the two rules mentioned on the previous page. The proof can be completed by showing that ϕ is a homeomorphism. See [Mo, Wi] for the full details. □

Theorem 6.10. *For the two-dimensional coin model, if a collision of m_{L} or m_{R} is labeled by L or R respectively, then any infinite sequence of L's and R's can be realized by choosing an appropriate initial condition.*

Proof. The sequence set Σ contains all the infinite sequences of $\mathrm{L}_1, \mathrm{L}_2, \mathrm{L}_3, \mathrm{R}_1, \mathrm{R}_2, \mathrm{R}_3$'s which do not violate the rules. From the rules, we see that if we set the representatives $[\mathrm{L}] = \{\mathrm{L}_1, \mathrm{L}_2, \mathrm{L}_3\}$ and $[\mathrm{R}] = \{\mathrm{R}_1, \mathrm{R}_2, \mathrm{R}_3\}$, then either $[\mathrm{L}]$ or $[\mathrm{R}]$ may follow after $[\mathrm{L}]$ (see also Figure 6.6). Similarly, after $[\mathrm{R}]$, either $[\mathrm{L}]$ or $[\mathrm{R}]$ may follow. In other words, Σ contains all possible sequences of $[\mathrm{L}]$'s and $[\mathrm{R}]$'s.

By Theorem 6.9, having all sequences of $[\mathrm{L}]$'s and $[\mathrm{R}]$'s implies that for any collision sequence of L's and R's, there exists a phase point of the coin billiard which makes collisions according to the sequence.

□

CHAPTER 7

Three-period orbits in billiards on the surfaces of constant curvature

Recall from Chapter 2.3 the natural coordinates of the phase space of a billiard (q, α), and the notion of periodic orbits. In this chapter, we study the set of three-period orbits in billiards on the hyperbolic plane and on the 2-sphere. Our main theorems are

Theorem 7.1. *The set of three-period orbits in any billiard on the hyperbolic plane has a measure zero.*

Theorem 7.2. *Let P_3 be the set of three-period orbits in the billiard domain $\partial\mathcal{Q}$ on the 2-sphere. Assume that some $(x_0, x_1, x_2) \in P_3$ has perimeter $L = \pi, 3\pi$ or 5π and that sufficiently small pieces of $\partial\mathcal{Q}$ containing x_0, x_1, x_2 belong to great circles. Then $(x_0, x_1, x_2) \in int(P_3)$ and P_3 has positive measure. Otherwise, $(x_0, x_1, x_2) \notin int(P_3)$.*

To prove the theorems, we extend the Jacobi fields based approach in [Wo2] and present the unified proof which treats all three billiard systems on the constant curvature manifolds in the same manner. Our argument proceeds independently of the underlying geometry until we get the compatibility condition. Then, using the relevant cosine formula (which depends on the geometry), we obtain the relation that must be satisfied on an open set containing only three-period orbits

$$k_g(q_0) = \sin^3(\alpha_0)F(L),$$

where L is the length of three-period orbits, $k_g(q_0)$ is geodesic curvature at one of the vertices, and α_0 is the angle of the billiard orbit with tangent at this vertex. The function

$F(L)$ depends on the underlying Riemannian manifold

$$F(L) = \begin{cases} \frac{2}{L} & \text{on } \mathbb{R}^2 \\ \coth\left(\frac{L}{2}\right) & \text{on } \mathbb{H}^2 \\ \cot\left(\frac{L}{2}\right) & \text{on } S^2. \end{cases}$$

From this formula, it is possible to classify sets of three-period orbits.

Remark 7.3. We only prove that the set of three-period orbits on $\mathbb{H}^2$ (and on $\mathbb{S}^2$ when special condition is not satisfied) has an empty interior. The stronger statement about zero measure follows verbatim the argument in [Wo2].

7.1 Billiard system on the surface of constant curvature

7.1.1 Jacobi fields

Let $\mathcal{Q}$ be a smooth domain on a surface of constant curvature $\varkappa$. The billiard ball inside $\mathcal{Q}$ travels along the geodesics and reflects at the boundary. Let $\boldsymbol{\gamma}(\epsilon, \tau)$, be a two parameter family of geodesics where $|\epsilon| < \epsilon_0, -\infty < \tau < \infty$.

We briefly recall the derivation of the Jacobi fields, see *e.g.* [Ca] or [Bar]. The Jacobi field is defined by

$$\mathbf{J}(\tau) = \frac{\partial \boldsymbol{\gamma}(0, \tau)}{\partial \epsilon}.$$

and it satisfies the Jacobi equation

$$\frac{\nabla}{d\tau}\frac{\nabla}{d\tau}\mathbf{J}(\tau) + R(\mathbf{J}(\tau), \dot{\boldsymbol{\gamma}})\dot{\boldsymbol{\gamma}} = 0,$$

where ∇ denotes the covariant derivative and R is the curvature tensor. As usual, we are interested in the component of the Jacobi field, that is perpendicular to $\dot{\boldsymbol{\gamma}}$. Therefore, it can

be expressed as

$$\mathbf{J}(\tau) = J(\tau)\mathbf{n}(\tau),$$

where $J(\tau)$ is a scalar function and $\mathbf{n}(\tau)$ is a unit vector field perpendicular to $\dot{\boldsymbol{\gamma}}$. If the surface has constant curvature $\varkappa$, then one obtains a scalar equation with constant coefficients

$$J''(\tau) + \varkappa J(\tau) = 0 \tag{7.1}$$

According to the standard result in the theory of differential equations, the solution of the Jacobi equation is uniquely defined if two initial conditions $J(0)$ and $J'(0)$ are given.

7.1.2 Evolution and reflection matrices

Consider the billiards on the hyperbolic plane $\mathbb{H}^2$ and the 2-sphere $\mathbb{S}^2$ which have the curvature $\varkappa = -1$ and 1 respectively. Solving the Jacobi equation (7.1), we get

$$J(\tau) = \begin{cases} J(0)\cosh(\tau) + J'(0)\sinh(\tau) & \text{on } \mathbb{H}^2 \\ J(0)\cos(\tau) + J'(0)\sin(\tau) & \text{on } S^2 \end{cases}$$

In each case, we obtain the evolution matrix $P(\tau)$,

$$\begin{pmatrix} J(\tau) \\ J'(\tau) \end{pmatrix} = P(\tau) \begin{pmatrix} J(0) \\ J'(0) \end{pmatrix},$$

$$\text{where } P(\tau) = \begin{cases} \begin{pmatrix} \cosh(\tau) & \sinh(\tau) \\ \sinh(\tau) & \cosh(\tau) \end{pmatrix} & \text{on } \mathbb{H}^2 \\ \\ \begin{pmatrix} \cos(\tau) & \sin(\tau) \\ \sin(\tau) & \cos(\tau) \end{pmatrix} & \text{on } S^2 \end{cases}$$

which describes the changes of the Jacobi field over time.

Note that the corresponding evolution matrix in the Euclidean case is given by

$$P(\tau) = \begin{pmatrix} 1 & \tau \\ 0 & 1 \end{pmatrix}.$$

When the ball hits the boundary at $x = (q, \alpha)$, the Jacobi field is transformed by the linear map $R(x)$ which is essentially the same as the reflection map in the Euclidean case

$$\begin{pmatrix} J_{out} \\ J'_{out} \end{pmatrix} = R(x) \begin{pmatrix} J_{in} \\ J'_{in} \end{pmatrix},$$

$$\text{where } R(x) = \begin{pmatrix} -1 & 0 \\ \dfrac{2k_g(q)}{\sin(\alpha)} & -1 \end{pmatrix}.$$

Note that the reflection matrix is directly related to the classical mirror formula [Wo2].

One should expect that the reflection matrix $R(x)$ for the billiard on a two-dimensional Riemannian manifold should have the same form as in the Euclidean case [Wo2]. Nevertheless, we provide some justification. Consider a one-parameter family of geodesics $\boldsymbol{\gamma}(\epsilon, \tau)$ reflecting from the billiard boundary $\partial\mathcal{Q}$ on a two-dimensional Riemannian manifold. In an ϵ-neighborhood of the reflection point x_0 of $\boldsymbol{\gamma}(0, \tau_0)$, the manifold can be represented as a smooth two-dimensional surface in $\mathbb{R}^3$. Projecting the geodesics and the boundary onto the tangent plane at x_0, we obtain the corresponding structure on the plane: family of orbits reflecting from the boundary. It is easy to estimate that the angles before and after the projection, differ by It is easy to estimate that the angles as well as distances before and after the projection differ by $O(\epsilon^2)$. This is mainly due to the expansion $\cos\epsilon = 1 - \epsilon^2/2 + ...$ Also, straightforward estimates show that the projected boundary curve will have the curvature equal to the geodesic curvature of $\partial\mathcal{Q}$ with the accuracy $O(\epsilon^2)$. As quadratic terms do not affect linear transformations, the reflection map will have the same form as in the Euclidean case with k being replaced by k_g.

7.2 Billiard on the hyperbolic plane

Assume that there is an open set of three-period orbits. Then we must have T^3 and DT^3 equal to the identity, which implies

$$P(\tau_1)R(x_1)P(\tau_0)R(x_0)(P(\tau_2)R(x_2) = I \tag{7.2}$$

where I is the identity map; x_0, x_1 and x_2 as the collision points; and τ_0, τ_1, τ_2 are the distances between collision points (Figure 7.1). This relation can be also rewritten as

$$P(\tau_1)R(x_1)P(\tau_0) = R^{-1}(x_2)P^{-1}(\tau_2)R^{-1}(x_0), \tag{7.3}$$

which takes the form

$$\begin{bmatrix} \cosh(\tau_1) & \sinh(\tau_1) \\ \sinh(\tau_1) & \cosh(\tau_1) \end{bmatrix} \cdot \begin{bmatrix} -1 & 0 \\ \dfrac{2k_g(q_1)}{\sin(\alpha_1)} & -1 \end{bmatrix} \cdot \begin{bmatrix} \cosh(\tau_0) & \sinh(\tau_0) \\ \sinh(\tau_0) & \cosh(\tau_0) \end{bmatrix}$$

$$= \begin{bmatrix} -1 & 0 \\ \dfrac{2k_g(q_2)}{\sin(\alpha_2)} & -1 \end{bmatrix} \cdot \begin{bmatrix} \cosh(\tau_2) & \sinh(\tau_2) \\ \sinh(\tau_2) & \cosh(\tau_2) \end{bmatrix} \cdot \begin{bmatrix} -1 & 0 \\ \dfrac{2k_g(q_0)}{\sin(\alpha_0)} & -1 \end{bmatrix}.$$

After simplification, we equate the top right components to get

$$\sinh(\tau_0 + \tau_1) - \sinh(\tau_2) = \frac{2k_g(q_1)\sinh(\tau_0)\sinh(\tau_1)}{\sin(\alpha_1)}. \tag{7.4}$$

This is comparable to the corresponding formula in the Euclidean case:

$$\tau_0 + \tau_1 - \tau_2 = \frac{2k(q_1)\tau_0\tau_1}{\sin(\alpha_1)},$$

which was derived in [Wo2].

We define θ to be the interior angle between two adjacent segments of an orbit, that is

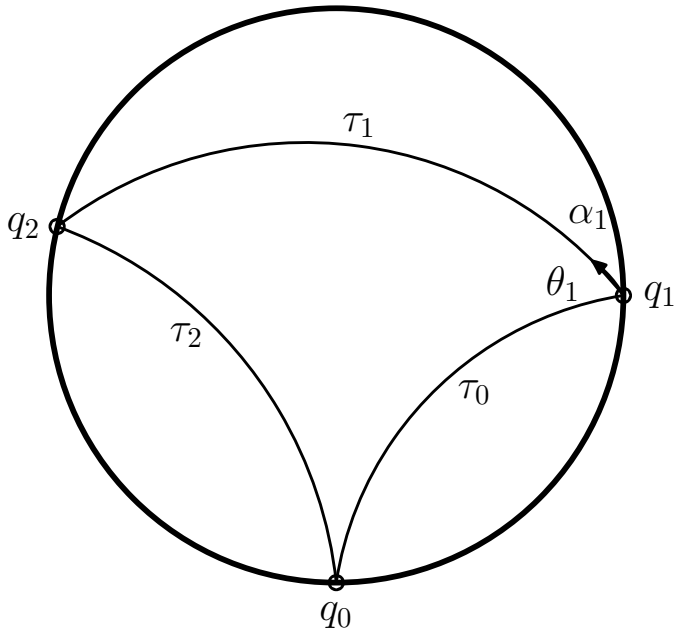


Figure 7.1: Three-period orbits on the hyperbolic plane

$\theta = \pi - 2\alpha$ (see Figure 7.1). Then we alter the hyperbolic cosine formula into

$$\cosh(\tau_2) = \cosh(\tau_0 + \tau_1) - \sinh(\tau_0)\sinh(\tau_1) - \sinh(\tau_0)\sinh(\tau_1)\cos(\theta_1).$$

We use the half angle formula to get

$$\cosh(\tau_0 + \tau_1) - \cosh(\tau_2) = 2\cos^2\left(\frac{\theta_1}{2}\right)\sinh(\tau_0)\sinh(\tau_1) \tag{7.5}$$

Combining (7.4) and (7.5), we arrive at

$$\frac{\cosh(\tau_0 + \tau_1) - \cosh(\tau_2)}{\cos^2\left(\frac{\theta_1}{2}\right)} = \frac{\sinh(\tau_0 + \tau_1) - \sinh(\tau_2)\cos\left(\frac{\theta_1}{2}\right)}{k_g(q_1)}.$$

Note that the length of an orbit $L = \tau_0 + \tau_1 + \tau_2$ is invariant. Therefore,

$$\begin{aligned} k_g(q_1) &= \frac{\sinh(\tau_0+\tau_1)-\sinh(\tau_2)}{\cosh(\tau_0+\tau_1)-\cosh(\tau_2)}\cos^3\left(\frac{\theta_1}{2}\right) \\ &= \frac{\sinh(L-\tau_2)-\sinh(\tau_2)}{\cosh(L-\tau_2)-\cosh(\tau_2)}\cos^3\left(\frac{\theta_1}{2}\right) \\ &= \cos^3\left(\frac{\theta_1}{2}\right)\coth\left(\frac{L}{2}\right) \\ &= \sin^3(\alpha_1)\coth\left(\frac{L}{2}\right) \end{aligned}$$

This relation must hold for all nearby orbits. In particular, for all orbits starting at the same point on the boundary with different angles of reflection. Thus, we obtain a contradiction because the right-hand side of the equation is not constant in any interval. Therefore, the set of three-period orbits has an empty interior. Next, following an argument in [Wo2], we obtain that the set has zero measure, which ends the proof of the Theorem 7.1.

7.3 Billiard on the 2-sphere

Now we prove Theorem 7.2 using the similar method. Assuming there is an open set of three-period orbits on S^2, we obtain again that T^3 and DT^3 are equal to the identity. Therefore, using (7.3) again

$$P(\tau_1)R(x_1)P(\tau_0) = R^{-1}(x_2)P^{-1}(\tau_2)R^{-1}(x_0),$$

we get

$$\sin(\tau_0)\cos(\tau_1)+\cos(\tau_0)\sin(\tau_1)-\sin(\tau_2) = \frac{2k_g(q_1)\sin(\tau_0)\sin(\tau_1)}{\sin(\alpha_1)}. \tag{7.6}$$

Note that this relation is the same as (7.4) if trigonometric functions are replaced with their hyperbolic counterparts (Figure 7.2).

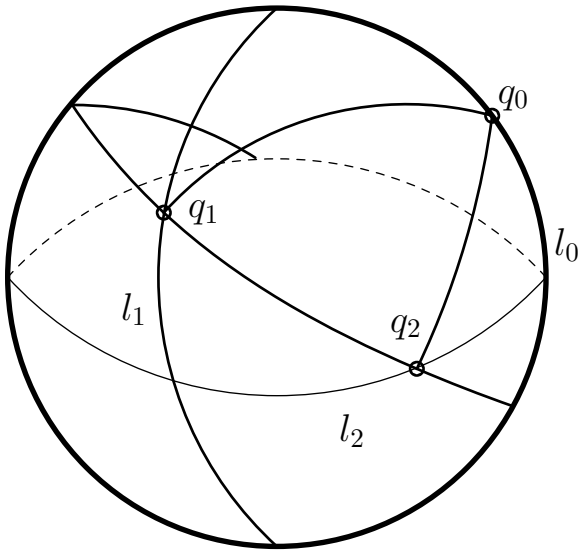


Figure 7.2: The special case $L = \pi$.

Combining (7.6) and the modified version of spherical cosine formula, we arrive at

$$k_g(q_1) = \sin^3(\alpha_1) \cot\left(\frac{L}{2}\right). \tag{7.7}$$

If $\cot\left(\frac{L}{2}\right) \neq 0$, then we have the same contradiction as in the hyperbolic case. When $L = (2n+1)\pi$, we have $\cot\left(\frac{L}{2}\right) = 0$ and $k_g = 0$. In this case, there could exist open sets of three-period orbits.

Now, we discuss the characteristics of the billiards on which open sets of three-period orbits exist. Note that only orbits without repetition are considered, and this assumption limits our cases to $L = \pi, 3\pi$, or 5π, see also [Gu].

Proposition 7.4. *Consider a spherical triangle $\triangle q_0 q_1 q_2$ on the unit sphere with perimeter $L = \pi, 3\pi$ or 5π. Let l_0, l_1, l_2 be the great circles passing through the vertices orthogonal to the corresponding bisectors. Then, these great circles intersect at the right angles and any billiard boundary containing segments of l_0, l_1, l_2 passing through q_0, q_1, q_2 will have an open set of three-period orbits.*

Proof. Let l_0 be a geodesic on S^2 and q_0 be any point on l_0. Create two geodesics, l_1 and l_2, that are perpendicular to l_0 and to each other, but do not pass through q_0. Consider

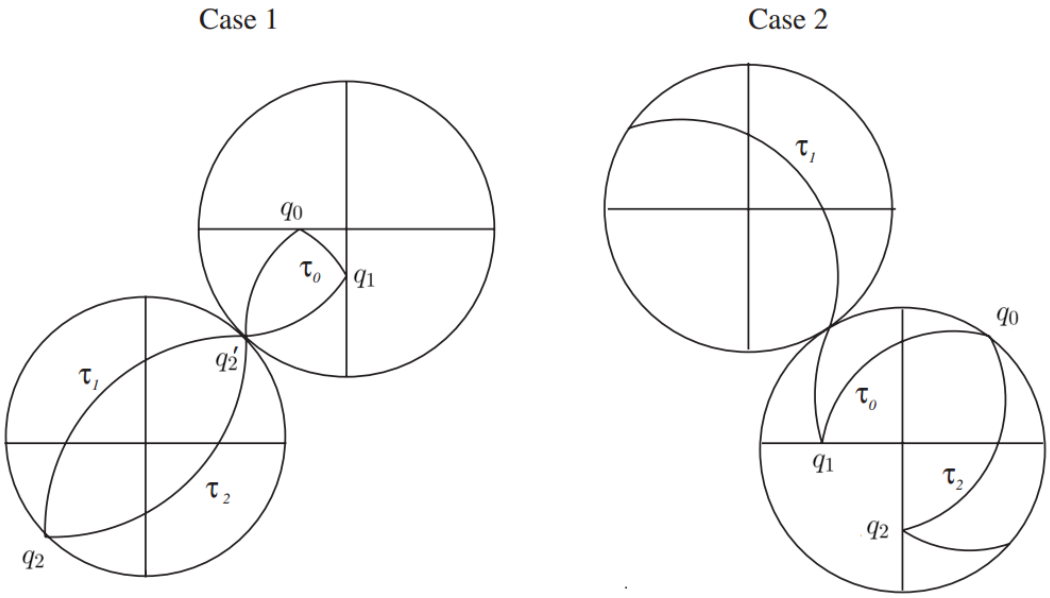


Figure 7.3: The special cases when the orbit length is $L = 3\pi$ or $L = 5\pi$.

any geodesic segment, τ, of length π, whose endpoint is q_0. Denote the angle between τ and l_0 as σ. Through two reflections over l_1 and l_2, this line τ forms a triangle of length π within the boundary created by l_0, l_1, and l_2. Since q_0 and σ were arbitrary, any three-period orbit of length π must be contained in one octant, which is formed by l_0, l_1, and l_2, whose intersections are orthogonal. In particular, this implies that all orbits in the octant are three-periodic except those which hit the corners.

Consider three great circles that intersect at q_0, q_1, and q_2. The total length of the lines is 6π. This implies that an orbit of length 5π is the complement of $\triangle q_0 q_1 q_2$. It follows that an orbit of length 5π must have vertices on l_0, l_1, l_2 as in the π case.

Now we consider a three-period orbit of length 3π. Since it is impossible to create an orbit where $\tau_0 = \tau_1 = \tau_2 = \pi$, we look at the two other possible cases; when $0 < \tau_0 < \pi$, $\pi < \tau_1, \tau_2 < 2\pi$, and $0 < \tau_0, \tau_1 < \pi$, $\pi < \tau_2 < 2\pi$. In case 1, as shown in Figure 7.3, we know that $\triangle q_0 q_1 q_2'$ has perimeter π and that q_2 is antipodal to q_2'. This implies that q_2 lies on l_2. Note that case 2 is simply the complement of case 1. Therefore, we conclude that a three-period orbit of length 3π also has vertices on l_0, l_1, and l_2. $\square$

The last proposition completely classifies the special cases when open sets of three-period orbits occur. If a given three-period orbit has perimeter $L \neq \pi, 3\pi, 5\pi$ then the relation (7.7) implies that this orbit has an empty interior in P_3. If $L = \pi, 3\pi$ or 5π but the geodesic curvature k_g does not vanish on the open arcs of the boundary passing through the vertices, then (7.7) again leads to the same contradiction.

Finally applying the argument in [Wo2], we obtain that if the special cases do not occur the set of three-period orbits has zero measure. This ends the proof of Theorem 7.2.

References

[Ar] V. I. Arnold, *Mathematical methods of classical mechanics*, Springer-Verlag, 1978.

[Bar] C. Bar, *Elementary Differential Geometry*, New York: Cambridge University Press, 2010.

[Ba] Yu. Baryshnikov and V. Zharnitsky, *Billiards and nonholonomic distributions*, J. Math. Sci. 128 (2005) 2706-2710.

[Ba2] Yu. Baryshnikov, V. Blumen, K. Kim, V. Zharnitsky, *Billiard dynamics of bouncing dumbbell*, Phys. D 269 (2014) 21-27.

[Ba3] Yu. Baryshnikov, *Spherical billiards with periodic orbits*, preprint.

[Bl] V. Blumen, K. Kim, J. Nance, V. Zharnitsky, *Three-period orbits in billiards on the surfaces of constant curvature*, Int. Math. Res. Not. 21 (2012) 5014-5024.

[Bu] D. Burago, S. Ferleger, A. Kononenko, *A geometric approach to semi-dispersing billiards* (survey), Ergodic Theory Dynam. Systems 18 (1998) 303-319.

[Ca] M. do Carmo, *Differential geometry of curves and surfaces*, New Jersey: Prentice-Hall Inc., 1976.

[Co] D. Cowan, *A billiard model for a gas of particles with rotation*, Discrete Contin. Dyn. Syst. 22 (2008) 101-109.

[Cv] P. Cvitanović, et al. *Chaos: Classical and Quantum*, `ChaosBook.org` (Niels Bohr Institute, Copenhagen 2016)

[Di] P. Diaconis, S. Holmes, R. Montgomery, *Dynamical bias in the coin toss*, SIAM Rev. 49-2 (2007) 211-235.

[Ge] D. Genin and S. Tabachnikov. *On configuration space of plane polygons, sub-Riemannian geometry and periodic orbits of outer billiards*, J. Mod. Dyn. 1-2 (2007), 155-173.

[Gl] A. A. Glutsyuk and Y. Kudryashov, *On quadrilateral orbits in planar billiards*, Dokl. Math. 83-3 (2011) 371-373.

[Gl2] A. A. Glutsyuk and Y. Kudryashov, *No planar billiard possesses an open set of quadrilateral trajectories*, J. Mod. Dyn. 6-3 (2012) 287-326.

[Gu] E. Gutkin and S. Tabachnikov, *Complexity of piecewise convex transformations in two dimensions, with applications to polygonal billiards on surfaces of constant curvature*, Mosc. Math. J. 6-4 (2006), 673-701, 772.

[Ho] P. J. Holmes, *The dynamics of repeated impacts with a sinusoidally vibrating table*, Journal of Sound and Vibration (1982) 84(2), 173-189.

[Iv] V. Ivrii, *The second term of the spectral asymptotics for a Laplace-Beltrami operator on manifolds with boundary*, Funct. Anal. Appl. 14-2 (1980) 98-106.

[Ke] J. B. Keller, *The probability of heads*, Amer. Math. Monthly 93-3 (1986) 191-197.

[Ki] K. Kim, *Chaotic dynamics of a bouncing coin*, preprint.

[Le] H. E. Lehtihet, B. N. Miller, *Numerical study of a billiard in a gravitational field*, Phys. D. 21 (1986) 93-104.

[Lv] M. Levi, *Qualitative Analysis of the Periodically Forced Relaxation Oscillations*, Mem. Amer. Math. Soc. (1981) 244.

[Ma] L. Mahadevan, E. H. Yong, *Probability, physics, and a coin toss*, Phys. Today (2011).

[Mi] T. Mizuguchi, M. Suwashita, *Dynamics of coin tossing*, Prog. Theor. Phys. Supplement 161 (2006) 274-277.

[Mo] J. Moser, *Stable and random motions in dynamical systems*, Princeton university press 77, 1973.

[Ry] M. Rychlik, *Periodic points of the billiard ball map in a convex domain*, J. Differential Geom. 30 (1989) 191-205.

[Si2] Ya. Sinai, *Introduction to Ergodic Theory*, in: Math. Notes, vol. 18, Princeton Univ. Press, Princeton, NJ, 1976.

[Si] Ya. Sinai, *Dynamical systems with elastic reflections: Ergodic properties of scattering billiards*, Russian Math. Surveys 25 (1970) 137-189.

[Sm] S. Smale, *Differentiable dynamical systems*, Bull. Amer. Math. Soc. 73-6 (1967) 747-817.

[Sto] L. Stojanov, *Note on the periodic points of the billiard*, J. Differential Geom. 34 (1991) 835-837.

[Str] J. Strzalko et al. *Dynamics of coin tossing is predictable*, Phys. Rep. 469 (2008) 59-92.

[Ta] S. Tabachnikov. *Billiards*, Panoramas et Synthèses 1, 1995.

[Ta2] S. Tabachnikov, *Geometry and Billiards*, in: Geometry and Billiards, vol. 30, AMS, 2005.

[Tu] A. Tumanov and V. Zharnitsky. *Periodic orbits in outer billiard*, Int. Math. Res. Not. 2006, Art. ID 67089, 17.

[Vo] Y. Vorobets, *On the measure of the set of periodic points of a billiard*, Math. Notes 55 (1994) 455-460.

[Vu] V. Z. Vulovic, R.E. Prange, *Randomness of a true coin toss*, Phys. Rev. A 33 (1986) 576-582.

[Wi] S. Wiggins, *Introduction to applied nonlinear dynamical systems and chaos*, Springer-Verlag, 1990.

[Wo] M. Wojtkowski, *A system of one dimensional balls with gravity*. Comm. Math. Phys. 126 (1990) 507-533.

[Wo2] M. Wojtkowski, *Two applications of Jacobi fields to the billiard ball problem*, J. Differential Geom. 40-1 (1994) 155-164.